Cahier du CRISCO
n°34

mai 2012

ISSN 2116-6242

Centre de Recherches Inter-langues sur la Signification en COntexte
E. A. 4255

POUR UNE LINGUISTIQUE ORIENTÉE OUTILS
La polysémie du verbe *compter* et les genres textuels

Jacques FRANÇOIS
&
Yacoub GHERISSI

CRISCO
Université de Caen (Bât. Sciences Porte SA S13), 14032 CAEN CEDEX
Tél. : 02 31 56 56 27 — Fax : 02 31 56 54 27 — Site web : www.crisco.unicaen.fr
Courriel direction : nicole.lequerler@unicaen.fr
Courriel secrétariat : liliane.docquiert@unicaen.fr

Le ***Centre de Recherches Interlangues sur la Signification en Contexte*** (CRISCO) contribue au débat linguistique à l'aide de trois outils éditoriaux :

I

Syntaxe & Sémantique

Revue de linguistique créée en 2000 aux Presses Universitaires de Caen

Responsables de la revue : J. FRANÇOIS & N. LE QUERLER

Comité de rédaction : Y.CAZAL, E. DELENTE, E. GILBERT, C. GUIMIER, I. HAIK, M. KRAUSE,P. LARRIVEE, D. LEGALLOIS, V. LENEPVEU, N. LE QUERLER, F. NEVEU, R. RENAULT.

Comité scientifique : P. BLUMENTHAL (Cologne), A. BORILLO (Toulouse-le-Mirail, Toulouse II), M. CHAROLLES (Sorbonne Nouvelle, Paris III), B. COMBETTES (Nancy II), J.P. DESCLES (Paris-Sorbonne, Paris IV), D. GEERAERTS (Louvain), B.N. GRUNIG (Paris VIII, Vincennes – Saint-Denis), P. HIRSCHBUHLER (Ottawa), G. KLEIBER (Sciences Humaines, Strasbourg II), P. KOCH (Tübingen), B. LAMIROY (Louvain), A. LEMARECHAL (Paris-Sorbonne, Paris IV), R. MARTIN (Paris-Sorbonne, Paris IV), S. MEJRI (Villetaneuse, Paris XIII), C. VET (Groningen), B. VICTORRI (CNRS, Paris).

n°1 ***Connecteurs et marqueurs de connexions***, sous la direction de Claude GUIMIER (mars 2000)

n°2 ***Sémantique du lexique verbal***, sous la direction de Françoise CORDIER, Jacques FRANÇOIS et Bernard VICTORRI (avril 2001)

n°3 ***Les grammaires du français et les "mots outils"***, sous la direction de Gudrun LEDEGEN et Nathalie ROSSI (mars 2002)

n°4 ***La valence vue d'Allemagne***, sous la direction de Peter BLUMENTHAL et Peter KOCH (mars 2003)

n°5 ***Polysémie et polylexicalité***, sous la direction de Salah MEJRI (avril 2004)

n°6 ***Aux marges de la prédication***, sous la direction de Irmtraud BEHR, Jacques FRANÇOIS, Anne LACHERET-DUJOUR & Florence LEFEUVRE (décembre 2005)

n°7 ***La terminologie linguistique – Problèmes épistémologiques, conceptuels et traductionnels***, sous la direction de Franck NEVEU (mars 2007)

n°8 ***Eléments de relation : de la phrase au texte***, sous la direction de Claude GUIMIER (décembre 2007)

n°9 ***Texte, documents numériques, corpus. Pour une science des textes instrumentée***, sous la direction de Mathieu VALETTE (décembre 2008)

n°10 ***Rhénania : dimensions syntaxiques et sémantiques***, sous la direction de Jean-Christophe PELLAT et Hélène VASSILIADOU (2010)

n°11 ***VARIA***, sous la direction de Jacques FRANÇOIS et Nicole LE QUERLER (2011)

n°12 ***VARIA***, sous la direction de Jacques FRANÇOIS et Nicole LE QUERLER (2012)

II

Bibliothèque de Syntaxe & Sémantique

collection de linguistique créée en 2002 aux Presses Universitaires de Caen

Traits d'union, sous la direction de Nicole LE QUERLER & Georges KLEIBER [Actes des *Journées de Linguistique Rhénane* de novembre 2000, organisées par SCOLIA et le CERLICO] (mai 2002)

L'adjectif en français et à travers les langues, sous la direction de Jacques FRANÇOIS [Actes du colloque de Caen, 28-30 juin 2001] (janvier 2005)

Composition syntaxique et figement lexical, sous la direction de Jacques FRANÇOIS & Salah MEJRI [Actes de l'atelier franco-tunisien de Caen, septembre 2002] (printemps 2006)

Autour de la préposition, sous la direction de Jacques FRANÇOIS, Eric GILBERT, Claude GUIMIER et Maxi KRAUSE (novembre 2009)

Jacques FRANÇOIS
Université de Caen-Basse-Normandie
CRISCO EA4255

Yacoub GHERISSI
Université de Carthage
Institut Supérieur des Langues de Tunis

POUR UNE LINGUISTIQUE ORIENTÉE OUTILS

La polysémie du verbe *compter* et les genres textuels

Ce 34e Cahier du CRISCO s'adresse en priorité aux étudiants de Master et de Doctorat en Sciences du langage à l'université de Caen et en Langue française à l'université de Carthage (Institut Supérieur des Langues de Tunis) et dans les autres institutions assurant une formation équivalente en Tunisie, qui s'interrogent sur l'utilité théorique et la faisabilité d'une analyse de corpus textuels.

Il prend la suite du Cahier n° 29 (Jacques François, 2010a, ***L'attestation des combinaisons lexicales à l'aide de la base de données textuelles FRANTEXT***), téléchargeable sur le site du CRISCO (www.crisco.unicaen.fr) et sur le site www.interlingua.fr. Nous aurons l'occasion dans ce nouveau cahier de renvoyer au Cahier 29 pour ce qui concerne l'exploitation de la base FRANTEXT.

Ce cahier profite de la collaboration entre Jacques François de l'université de Caen, fondateur du CRISCO, et Yacoub Ghérissi, spécialiste des corpus textuels à l'université de Carthage. Le Cahier se compose de deux parties :

- La première se présente comme un article de linguistique traditionnel en ce sens que trois corpus textuels du français contemporain, respectivement littéraire, journalistique et politique, y sont exploités et comparés selon les méthodes de l'analyse comparative de corpus. Les techniques permettant la constitution et l'exploitation des corpus textuels n'y sont pas décrites, pas plus qu'elles ne le sont habituellement dans les articles de linguistique destinés à une communauté supposée en avoir déjà une certaine connaissance.

- La seconde vise au contraire à entrer dans le détail de ces techniques, afin que les résultats présentés en première partie soient reproductibles. Il s'agit de décrire les outils – relativement élémentaires, en particulier le concordancier AntConc qui y occupe une place centrale – qui permettent de reconstituer pas à pas notre démarche (et éventuellement de dénicher nos erreurs).

Si l'on entend par « analyse de corpus textuels » une méthode capable de s'appliquer AUTOMATIQUEMENT (c'est-à-dire sans l'intervention du linguiste, une

fois que la machine est en marche) et EN AVEUGLE, c'est-à-dire destinée à dégager une liste de constructions distinctes à partir d'une description rigoureusement distributionnelle, nous ne pratiquons pas ce type d'analyse de corpus.

Notre conviction, confortée par le développement des « grammaires de construction » (voir le Cahier 26 du CRISCO, Jacques François, 2008, ***Les grammaires de construction – Un bâtiment ouvert aux quatre vents***), est que chaque CONSTRUCTION est (dans la terminologie de Ronald Langacker) une ***assemblée symbolique*** entre une expression (une structure morpho-syntaxique) et une signification (grossièrement approchée à l'aide d'une paraphrase) et que l'analyse des constructions d'un verbe figurant dans différents corpus doit intégrer d'une part une analyse distributionnelle et transformationnelle dans le sens de Zellig Harris et Maurice Gross, et d'autre part une analyse sémantique.

C'est ce que nous avons cherché à faire ici à propos du verbe ***compter***, sélectionné en raison du large éventail de CONSTRUCTIONS dans lesquelles il peut figurer. Nous chercherons à montrer que la polysémie virtuelle d'un verbe comme ***compter*** (dont nous distinguons 13 CONSTRUCTIONS) est en fait réduite (mais non levée) dans chacun des trois corpus littéraire, journalistique et politique que nous avons examinés et que chaque réduction de la polysémie a un profil particulier. Il s'agit donc ici de notre part d'un **PLAIDOYER POUR UNE PRATIQUE COMPARATIVE ET QUANTITATIVE DE L'ANALYSE DES CORPUS TEXTUELS**.

SOMMAIRE

page

I. CE QUE DEVIENT LA POLYSÉMIE DU VERBE *compter* DANS TROIS CORPUS LITTÉRAIRE, JOURNALISTIQUE ET POLITIQUE 4

I-1. Deux articles lexicographiques en contraste 4

I-2. Le dictionnaire en ligne DicoValence 7

I-3. Le profil d'emploi du verbe **compter** *dans le corpus littéraire* 12

I-4. Le profil d'emploi du verbe **compter** *dans le corpus journalistique* 16

I-5. Le profil d'emploi du verbe **compter** *dans le corpus des discours politiques* 19

I-6. Le profil comparé des emplois du verbe compter dans les trois corpus 21

I-7. Conclusion intermédiaire 24

II. TROIS OUTILS ÉLÉMENTAIRES DE LINGUISTIQUE QUANTITATIVE 26

II-1. Les bases textuelles 26

II-1.1. Sites de recherche en ligne 26

II-1.2. Sites de téléchargement de textes numérisés 28

II–2. Concordancier et éditeurs de textes 30

II–2.1. Le concordancier AntConc 30

II–2.2. WORD de Microsoft Office 39

II–2.3. EXCEL de Microsoft Office 41

II–2.4. Combinaison AntConc / EXCEL 48

II–2.5. Conclusion 55

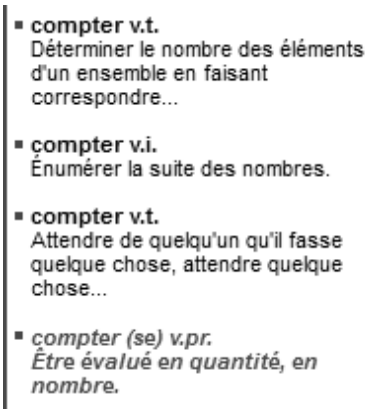
RÉFÉRENCES 56

1. CE QUE DEVIENT LA POLYSÉMIE DU VERBE *compter* DANS TROIS CORPUS LITTÉRAIRE, JOURNALISTIQUE ET POLITIQUE

1.1. Les articles *compter* dans deux dictionnaires d'usage

Avant de chercher à évaluer le poids des différents types d'emploi du verbe *compter* dans trois corpus textuels différents, il est nécessaire de mesurer le degré d'accord entre les dictionnaires sur le dénombrement et la description de ces emplois. Pour cela, nous commencerons par comparer deux dictionnaires d'usage très large, d'une part le *Petit Larousse* disponible en ligne (► http://www.larousse.fr/dictionnaires/francais/) et d'autre part le *Petit Robert* (édition électronique de 2011).

Le *Petit Larousse* en ligne dégroupe ces emplois en quatre articles, le premier regroupant les emplois transitifs avec 13 entrées, le second les emplois intransitifs (ou pour certains, transitifs indirects, selon le statut attribué au complément prépositionnel, par ex. *parmi N<coll>*) avec 6 entrées, le troisième des emplois transitifs indirects et le quatrième l'emploi pronominal à valeur générique.


▪ **compter v.t.**
Déterminer le nombre des éléments d'un ensemble en faisant correspondre...

▪ **compter v.i.**
Énumérer la suite des nombres.

▪ **compter v.t.**
Attendre de quelqu'un qu'il fasse quelque chose, attendre quelque chose...

▪ *compter (se) v.pr.*
Être évalué en quantité, en nombre.

Le statut des compléments seconds prépositionnels n'est pas clairement distingué de celui des circonstants, en particulier en ce qui concerne la préposition *pour*, introduisant un complément en (1) et un circonstant en (2-3) :

(1) *Vous pouvez compter Pierre **pour** trois personnes, il mange comme trois.* ► Entrée trans.7
(2) *Il m'a compté 100 euros (656 francs) **pour** la réparation*. ► Entrée trans.4
(3) *Il faut compter six bonnes heures **pour** faire ce travail.* ►Entrée trans.6

La troisième colonne du tableau 1 ci-dessous indique les entrées correspondantes du *Petit Robert*. On constate que toutes les entrées du *Petit Larousse* ont un pendant dans le *Petit Robert*, mais dans un ordre différent.

Larousse en ligne[1]	**L⇒P**
V. trans.	
1. Déterminer le nombre des éléments d'un ensemble en faisant correspondre à chacun d'entre eux, et successivement, chacun des nombres de la série des entiers naturels ; dénombrer : *Compter les élèves de la classe.*	I.1
2. Déterminer le nombre, la quantité de choses, de personnes en procédant à un calcul : *N'oubliez pas de compter vos heures pour pouvoir vous faire payer.*	I.1
3. Faire entrer quelqu'un, quelque chose dans un calcul d'ensemble : *Il vous faut trois mètres de tissu en comptant l'ourlet.*	I.5
4. Évaluer le prix de quelque chose et se le faire payer : *Il m'a compté 100 euros (656 francs) pour la réparation.*	I.2
5. Verser une somme à quelqu'un, la lui payer : *Passez à la caisse, le caissier vous comptera vos 100 €.*	I.2
6. Évaluer une certaine quantité à l'avance, la prévoir : *Il faut compter six bonnes heures pour faire ce travail.*	I.3
7. Attribuer telle valeur à quelqu'un, quelque chose : *Vous pouvez compter Pierre pour trois personnes, il mange comme trois.*	I.5 ?
8. Comporter tel nombre de choses, de personnes, en être constitué, l'avoir, le comprendre : *Collection qui compte deux mille gravures.*	I.7
9. Avoir pour soi des choses ou des personnes au sein d'un ensemble : *Il compte quelques personnalités de la finance parmi ses amis.*	I.6
10. Avoir tel nombre d'années à son actif, avoir une durée de tant : *Il compte 20 ans de présence dans la maison.*	I.4
11. Prendre quelque chose en considération, en tenir compte : *Et je ne compte pas la fatigue, les heures d'angoisse.*	I.3
12. Donner quelque chose à quelqu'un avec parcimonie : *On a l'impression qu'il nous compte tout ce qu'il offre.*	I.2
13. Dénombrer les moments qui s'écoulent ou qui sont à venir en trouvant le temps long : *Compter les jours qui séparent des vacances.*	I.3
V. intrans	
1. Énumérer la suite des nombres : *Enfant qui sait compter jusqu'à 10.*	II.1
2. Faire des calculs : *Vous vous êtes trompé dans votre addition, apprenez à compter.*	II.1
3. Calculer avec précision ses dépenses par souci d'économie : *Avec un budget aussi restreint, il faut sans cesse compter.*	II.1
4. Entrer en ligne de compte avec telle valeur : *C'est une erreur, ça ne compte pas.*	II.5
5. Avoir de l'importance : *Pierre compte beaucoup pour elle.*	II.4
6. Être au rang, au nombre de : *Livre qui compte parmi les plus grands succès de l'année.*	II.6
V. trans.ind.	
1. Attendre de quelqu'un qu'il fasse quelque chose, attendre quelque chose de quelque chose, mettre son espoir en quelqu'un ou en quelque chose : *Je compte sur vous pour m'aider.*	II.3
2. Avoir l'intention de, se proposer de : *Vous comptez venir ?*	I.10
V.pron.	
Être évalué en quantité, en nombre : *Ses exploits se comptent sur les doigts de la main.*	III

Tableau 1 : Les quatre articles *compter* du *PL* en ligne

Le PR ne pratique pas de dégroupement et distribue son article *compter* en trois rubriques : emplois transitifs, intransitifs et pronominal. La première compte 10 entrées (dont une classée '*vieilli*' c'est-à-dire sortie de l'usage actuel), la seconde 7 entrées (dont une '*vx*.', c'est-à-dire archaïque et généralement inconnue des usagers) et la troisième une seule. La comparaison entre les

[1] Les rubriques originales du Larousse en ligne ne sont pas numérotées. Le dictionnaire dégroupe le verbe en deux articles (transitif vs. intransitif)

microstructures de l'article du PR et des deux articles du PL (colonne de droite du tableau 2) révèle, contrairement à la colonne de droite du tableau 1, que certaines entrées du PR n'ont pas de pendant dans le PL. Elles figurent en grisé sans le tableau 2, outre les deux entrées classées comme 'vieilli' ou 'vx.' Mais elles ne sont qu'au nombre de 3, ce qui suggère que la couverture des types d'emploi (et de constructions) est légèrement meilleure dans le PR que dans le PL. Cette constatation n'est pas étonnante dans la mesure où les deux dictionnaires visent un lectorat différent : le PR s'adresse à des usagers généralement lettrés et cultivés et le PL plutôt aux élèves de l'enseignement élémentaire et aux collégiens.

Petit Robert électronique[2]	**P⇒L**
I. Verbe transitif	
1. (milieu XII^e^) Déterminer (une quantité) par le calcul ; spécialt Établir le nombre de.	1
2. Mesurer avec parcimonie.	4,5,12
Par ext. *Compter une somme à qqn,* la lui payer.	
3. Mesurer (le temps).	
4. Avoir duré (un certain temps).	6, 11
5. Prévoir, évaluer (une quantité, une durée).	10
6. Comprendre dans un compte, un total, une énumération.	3, 7
◆ *Compter parmi, au nombre de :* ranger au nombre de.	9
7. Comporter.	8
8. (début XIV^e^) **Vieilli** *Compter* (qqch.) *pour.*	**Ø**
9. Espérer, penser.	tr.ind.2
(1835) *Sans compter que :* sans considérer que.	**Ø**
II. Verbe intransitif	
1. Calculer.	1,2
2. **COMPTER AVEC, SANS (qqn, qqch.) :** tenir, ne pas tenir compte de.	**Ø**
3. (fin XVII^e^) **COMPTER SUR :** faire fond, s'appuyer sur.	tr.ind.1
4. (1863) Avoir de l'importance.	5
5. Être compté, évalué.	4
6. Être (parmi).	6
7. (milieu XIV^e^) **Vx → dater.**	**Ø**
III. Verbe pronominal SE COMPTER (réfléchi) Se mettre au nombre de.	**Ø**

Tableau 2 : L'article *compter* du *Petit Robert*, édition électronique 2011

[2] Nous ne reproduisons que le plan de l'article.

1.2. Le dictionnaire en ligne *DicoValence*

De la comparaison du PL et du PR pour les articles du verbe *compter*, il ressort un certain degré de désaccord entre les deux traitements lexicographiques affectant la **MACROSTRUCTURE** (deux articles dans le PL, un seul dans le PR) et la **MICROSTRUCTURE** (une rubrique pronominale dans le PR et quatre autres entrées absentes du PL, outre les deux entrées vieillie et archaïque).

Afin d'y voir plus clair, il est souhaitable de profiter d'un troisième dictionnaire supposé plus précis qui joue le rôle de '*tertium comparationis*', c'est-à-dire de 'tiers comparatif', propre à trancher sur la validité descriptive des deux descriptions du PL et du PR. Pour ce faire le *DicoValence* mis en ligne[3] par la *Katholieke Universiteit Leuven*[4] est particulièrement adapté. C'est un dictionnaire électronique spécifiquement dédié à la description lexicographique de la valence sémantaxique des verbes du français.

L'éventail des emplois du verbe *compter* (ce que nous appellerons désormais sa **POLYTAXIE**) est décrit en 17 entrées représentant 16 emplois prédicatifs et 1 emploi de verbe adjoint (apparenté aux 'coverbes' de M. Herslund, 2000 et aux verbes 'opérateurs' de M. Gross, 1975). Le tableau 3 reproduit les quatre premiers champs de cette description qui en compte plusieurs autres (évoqués dans le tableau 4). Le premier introduit la valence syntaxique, le second le type fonctionnel, le troisième fournit un exemple et le quatrième, le plus intéressant, le cadre sémantaxique, c'est-à-dire la sous-catégorisation de chaque actant[5].

	VAL	VTYPE	EG	FRAME
1	***compter: P0 P1 PP<pour>***	predicator simple	***je compte deux heures pour réparer votre machine à laver***	subj : pron|n : [hum], obj : pron|n : [abs], objp<pour> : pron|n|compl|inf : [abs,mood : subj]
2	***compter: P0 P1***	predicator simple	***le projet compte trois volets***	subj : pron|n : [abs], obj : pron|n : [abs]
3	***compter: P0 (P1)***	predicator simple	***il compte les arbres***	subj : pron|n : [hum], ?obj : pron|n : [hum,+complex]
4	***compter: P0 P1 PP<parmi>***	predicator simple	***elle te compte parmi ses confidentes***	subj : pron|n : [hum], obj : pron|n : [hum], objp<parmi> : pron|n : [hum,+complex]
5	***compter: P0 P1 PP<parmi>***	predicator simple	***je compte ce vin parmi les meilleurs de ma cave***	subj : pron|n : [hum], obj : pron|n : [nhum], objp<parmi> : pron|n : []

[3] Adresse internet : *bach.arts.kuleuven.be/**dicovalence/***

[4] Cette université néerlandophone est à distinguer de l'*Université Catholique de Louvain*, francophone, sise à Louvain-la-Neuve.

[5] Les entrées du type "predicator simple" sont prédicatives, celles du type "adjunct verb" ont un coverbe comme tête (la classe est mentionnée entre parenthèses). Les symboles argumentaux sont : P0 : paradigme valenciel sujet; P1 : paradigme objet direct ; PP<prep> : paradigme objet prépositionnel; PQ : paradigme de quantité; PMi: paradigme attribut de l'objet introduit par COMME (cf. van den Einde & Mertens, 2006).

6	***compter: P0 PP<parmi>***	predicator simple	***il compte parmi les plus grands pianistes du vingtième siècle***	subj : pron|n : [hum], objp<parmi> : pron|n : []
7	***compter: P0 P1 PP<dans>***	predicator simple	***je compte la communication téléphonique dans l'addition***	subj : pron|n : [hum], obj : pron|n : [nhum,?abs], objp<dans> : pron|n : [?nhum,abs]
8	***compter: P0 (PP<pour>)***	predicator simple	***je ne compte plus pour lui***	subj : pron|n : [hum,nhum,?abs], ?objp<pour> : pron|n : [hum]
9	***compter: P0 PP<pour>***	predicator simple	***mes remarques ne comptent pour rien, elles comptent pour du beurre***	subj : pron|n : [hum,nhum,?abs], objp<pour> : pron|n : []
10	***compter: P0 P1 PP<pour>***	predicator simple	***Cocteau comptait cette galerie pour une mi-salle de billard, un mi-cabinet de travail***	subj : pron|n : [hum], obj : pron|n : [hum], objp<pour> : pron|n : [?nhum,abs]
11	***compter: P0 PP<sur>***	predicator simple	***tout le monde compte sur toi***	subj : pron|n : [hum], **objp<sur> : pron|n : [hum,?nhum,abs]**
12	***compter: P0 PP<sur>***	predicator simple	***je compte qu'il sera seul***	subj : pron|n : [hum], **objp<sur> : pron|n|compl|inf : [abs,mood : ind/subj]**
13	***compter: P0 PQ (P2)***	predicator simple	***je lui ai compté trois salaires***	subj : pron|n : [hum], ?objà : pron|n : [hum], quant<> : pron|n : []
14	***compter: AdjunctVerb***	adjunct_verb: 6	***je compte rester deux jours de plus***	
15	***compter: P0 PP<avec, sans>***	predicator simple	***il faudra compter avec eux***	subj : pron|n : [hum], objp<avec,sans> : pron|n : [hum,?nhum,abs]
16	***compter: P0 P2 P1 PMi***	predicator simple	***il m'a compté cette remarque comme point d'honneur***	subj : pron|n : [hum], obj : pron|n : [abs], objà : pron|n : [hum], attr_obj<comme>
17	***compter: P0 PMi***	predicator simple	***ton numéro d'inscription compte comme bulletin de loterie***	subj : pron|n : [abs], attr_subj<comme>

Tableau 3 : Les 17 entrées de l'article *compter* de *DicoValence* (université de Louvain)

Cette description introduit implicitement la notion d'**ALTERNANCE** entre constructions syntaxiques développée dans deux cadres théoriques différents par B. Levin (1993) pour l'anglais et par J. Dubois & F. Dubois-Charlier (1997) pour le français. En effet les entrées 11 et 12 (notre numérotation) ont un même descriptif dans la colonne VAL(ence) comportant la mention PP<sur>, mais les réalisations de cette construction incluent en 12 *qn compte que P*, *qn compte INF* et leur pronominalisation *qn compte sur cela/ça/(là-)dessus*, alors que l'entrée

11 spécifie le complément de la préposition *sur* comme un animé[6] (voir les parties grisées du descriptif sémantaxique en colonne FRAME), cf. *Tout le monde compte sur toi*.

Plus mystérieuse est la distinction entre les entrées 4 et 5. En quoi peut exactement consister la différence entre les exemples des entrées 4 et 5 ?

(entrée 4) *Elle te compte parmi ses confidentes.*

(entrée 5) *Je compte ce vin parmi les meilleurs de ma cave.*

La prise en compte des champs suivants du descriptif original de *DicoValence* se révèle nécessaire (cf. tableau 4) :

Champ	Entrée 17410 (4)	Entrée 17415 (5)
VAL	***compter: P0 P1 PP<parmi>***	***compter: P0 P1 PP<parmi>***
VTYPE	predicator simple	predicator simple
VERB	COMPTER/compter	COMPTER/compter
NUM	17410	17415
EG	***elle te compte parmi ses confidentes***	***je compte ce vin parmi les meilleurs de ma cave***
TR_DU	rekenen (tot)	rekenen (tot)
TR_EN	**count (as)** , rate (among)	rate (among)
FRAME	subj:pron|n:[hum], **obj:pron|n:[hum]** , **objp<parmi>: pron|n: [hum,+complex]**	subj:pron|n:[hum], **obj:pron|n:[nhum**] , **objp<parmi>:pron|n:[]**
P0	qui, je, nous, elle, il, ils, on, celui-ci, ceux-ci	qui, je, nous, elle, il, ils, on, celui-ci, ceux-ci
P1	qui, te, vous, la, le, les, se réfl., se réc., en Q, celui-ci, ceux-ci, l'un l'autre	que, la, le, les, en Q, celui-ci, ceux-ci
PP_PR	parmi	parmi
PP	lesquels, **eux**, ceux-ci	lesquels, ceux-ci
RP	passif être, **se faire passif**	passif être
LC	17410-17420 il la compte parmi celles-là, elle compte parmi celles-là	17415-17420 je le compte parmi ceux-là, il compte parmi ceux-là
AUX	avoir	avoir

Tableau 4 : Les champs du descriptif des verbes dans DicoValence

On constate en mettant les entrées 4 et 5 en vis-à-vis trois traits distinctifs :

a) l'entrée 4 peut se traduire en anglais par *count (as),* la 5 seulement par *rate (among)* ;

[6] Le descriptif de ce complément est plus exactement [hum, ?nhum, abs]. Il faudrait savoir si « nhum » symbolise un animé non humain et quel exemple d'abs(trait) les concepteurs ont à l'esprit.

b) le cadre sémantaxique de l'entrée 4 spécifie l'objet premier comme [hum] contre [nhum] pour l'entrée 5, et l'objet second introduit par *parmi* comme [hum, +complex], contre [] pour l'entrée 5, c'est-à-dire une absence de spécification ;

c) la pronominalisation du PP n'inclut *parmi eux* que dans l'entrée 4.

On voit donc que la distinction est très subtile. La 'complexité' supposée de l'objet second de l'entrée 4 (illustré par *ses confidentes*) et implicitement l'absence de 'complexité' de celui de l'entrée 5 (*les meilleurs de ma cave*) laissent perplexe, et dans les deux cas le pendant intransitif (sémantiquement, une évaluation avec l'évaluateur/-trice introduit par *pour*) a une forme similaire (décrite dans l'entrée 6) :

[Entrée 4] *Elle te compte parmi ses confidentes.*
→ *Tu comptes parmi ses confidentes (pour elle)*

[Entrée 5] *Je compte ce vin parmi les meilleurs de ma cave.*
→ *Ce vin compte parmi les meilleurs de ma cave (pour moi).*

On peut donc considérer cette distinction comme superflue. Sur la base de l'article compter de *DicoValence*, nous pouvons cependant mieux évaluer comparativement les articles du PL et du PR. Le tableau 5 introduit la numérotation, la valence et l'exemple de chaque entrée du *Dicovalence* au milieu, à gauche la numérotation de l'entrée correspondante du PL et à droite celle de l'entrée correspondante du PR.

Larousse			**DicoValence**	**Exemple**		**PRE**
Tr.6	⇔	**1**	***compter: P0 P1 PP<pour>***	*je compte deux heures pour réparer votre machine à laver*	⇔	**I.5**
Tr.8	⇔	**2**	***compter: P0 P1***	*le projet compte trois volets*	⇔	**I.7**
Tr.1	⇔	**3**	***compter: P0 (P1)***	*il compte les arbres*	⇔	**I.1**
Tr.9	⇔	**4**	***compter: P0 P1 PP<parmi>***	*elle te compte parmi ses confidentes*	⇔	**I.9**
	⇔	**5**	***compter: P0 P1 PP<parmi>***	*je compte ce vin parmi les meilleurs de ma cave*	⇔	
Intr.6	⇔	**6**	***compter: P0 PP<parmi>***	*il compte parmi les plus grands pianistes du vingtième siècle*	⇔	**II.5**
Intr.3 ?	⇔	**7**	***compter: P0 P1 PP<dans>***	*je compte la communication téléphonique dans l'addition*	⇔	**I.6**
Intr.5	⇔	**8**	***compter: P0 (PP<pour>)***	*je ne compte plus pour lui*	⇔	**I.4**
Intr.4	⇔	**9**	***compter: P0 PP<pour>***	*mes remarques ne comptent pour rien, elles comptent pour du beurre*	⇔	**II.4**
Intr.7	⇔	**10**	***compter: P0 P1 PP<pour>***	*Cocteau comptait cette galerie pour une mi-salle de billard, un mi-cabinet de travail*	⇔	**I.6 ?**

Ø	⇔	**11**	***compter: P0 PP<sur>***	*tout le monde compte sur toi*	⇔	**II.3**
Ø	⇔	**12**	***compter: P0 PP<sur>***	*je compte qu'il sera seul*	⇔	**I.9**
I.4	⇔	**13**	***compter: P0 PQ (P2)***	*je lui ai compté trois salaires*	⇔	**I.2**
Ø	⇔	**14**	***compter: | AdjunctVerb***	*je compte rester deux jours de plus*	⇔	**I.9**
Ø	⇔	**15**	***compter: P0 PP<avec, sans>***	*il faudra compter avec eux*	⇔	**II.2**
Ø	⇔	**16**	***compter: P0 P2 P1 PMi***	*il m'a compté cette remarque comme point d'honneur*	⇔	**Ø**
Ø	⇔	**17**	***compter: P0 PMi***	*ton numéro d'inscription compte comme bulletin de loterie*	⇔	**Ø**

Tableau 5 : Comparaison des entrées de compter dans le PL et le PR par l'intermédiaire de *DicoValence*

Cette disposition permet de faire deux constatations :

a) Deux entrées de *DicoValence* n'ont de pendant dans aucun des deux autres dictionnaires : la 16 et la 17 qui comportent un attribut de l'objet introduit par *comme*. Le *Grand Robert* tout comme le *Trésor de la Langue Française* ignorent également ces constructions. Elles posent la question du statut du constituant introduit par *pour* dans les entrées 9 (*Mes remarques ne comptent **pour** rien, elles comptent pour du beurre*) et 10 (*Cocteau comptait cette galerie **pour** une mi-salle de billard, un mi-cabinet de travail*) par opposition au complément introduit par *pour* dans l'entrée 8 (*Je ne compte plus **pour** lui*). En effet en 8, *lui* désigne un expérient (8 ≈ *il estime que je ne compte plus*), tandis qu'en 9 et 10 nous sommes en présence de l'assimilation d'une entité à une autre entité (9 ≈ *Mes remarques **sont** pour lui **comme** du beurre* / 10 ≈ *Cette galerie **est** pour Cocteau **comme** une mi-salle de billard, un mi-cabinet de travail*). Et donc le **PP<pour>** y aurait un statut d'attribut de l'objet (en 10) ou du sujet (en 9) au même titre que le constitutant introduit par *comme* en 16 et 17.

b) Quatre entrées de *DicoValence* ont un pendant dans le PR mais pas dans le PL. Ce sont les entrées 11, 12, 14 et 15. L'évaluation comparative engagée dans le tableau 1, concernant l'absence de *pendant* dans le PL pour les entrées I.9-10 et II.2-3 du PR est ainsi confortée par le *DicoValence* exploité comme 'tiers comparatif' entre les descriptions lexicographiques du PL et du PR.

1.3. Le profil des emplois du verbe *compter* dans le corpus littéraire

Notre propos est de mettre en vis-à-vis le profil des emplois du verbe ***compter*** dans trois corpus, littéraire, journalistique et politique[7]. Nous avons souhaité que les trois corpus regroupent des écrits approximativement contemporains. Il s'agira pour le corpus littéraire de textes romanesques et poétiques de la dernière décennie du 20e siècle enregistrés dans la base FRANTEXT catégorisée, pour le corpus journalistique du mois de janvier 2002 de l'Est Républicain[8] et pour le corpus politique des discours regroupé dans la base de Jean Véronis à l'université d'Aix-Marseille[9]. Pour chaque corpus nous avons retenu aléatoirement[10] 200 occurrences. L'examen porte donc au total sur 600 occurrences.

Pour le corpus romanesque, nous avons opté pour la version catégorisée de la base de données textuelles FRANTEXT en raison du fort '*bruit*'[11] produit par la requête « toute forme du verbe ***compter*** ». Les textes les plus récents qui ont fait l'objet d'une catégorisation datent de 1997. Le corpus de 600 occurrences couvre la période 1980-97.

Sur la base de la comparaison des constructions figurant dans les trois corpus, nous avons établi une liste de 13 constructions. Les occurrences à classer se présentent sous la forme suivante (Tableau 6) :

[7] Pour la méthodologie d'analyse distributionnelle des occurrences constituant le corpus littéraire, voir la partie II.
[8] Les années 1999 et 2002 de l'Est Républicain sont téléchargeables sur le site du CNRTLà Nancy ► http://www.cnrtl.fr/corpus/estrepublicain/ .
[9] Ces discours sont téléchargeable sur demande adressée à Jean Véronis. On peut effectuer des requêtes lexicales sur le site ► http://www.veronis.fr/discours
[10] La sélection est aléatoire pour le corpus littéraire, car elle est déterminée uniquement par l'ordre chronologique inversé des occurrences en commençant par l'année 1997 et pour le corpus politique, car elle couvre un mois de l'année 2002 choisi au hasard. Pour le corpus politique qui est moins volumineux, les 1929 discours disponibles ont été pris en compte.
[11] Le 'bruit' désigne la proportion d'occurrences retenues par erreur par le logiciel de sélection. Dans le cas du verbe compter il s'agit essentiellement de la forme nominale ***compte*** (accessoirement aussi de la forme ***comptes***) classée par erreur comme verbale.

Références	pages	Cotexte gauche	*COMPTER*	Cotexte droit
ROLIN.J / L'ORGANI-SATION / 1996	pp. 105-106	La plupart des émeutiers étaient très jeunes. On	comptait	dans leurs rangs une proportion plus que satisfaisante de gamines au visage constellé de taches de rousseur, au charme desquelles il était difficile de résister (…)
SALVAYRE.L / LA PUISSANCE DES MOUCHES / 1995	pp. 93-94	Non, rassurez-vous, je ne le suis que dans les rêves. Et dans les rêves, à ce qu'il me semble, ces choses-là ne	comptent	pas. Mais depuis que j'existe, personne, en dehors de maman, ne m'a accordé la moindre attention, (…)
MAKINE.A / LE TESTAMENT FRANÇAIS / 1995	Pp. 111-112	Sa vie extravagante n'était consacrée qu'à l'amour. Et autour de ce trône s'agitaient des hommes - les uns	comptaient	les maigres napoléons de leur fortune anéantie, les autres approchaient lentement le canon de leur revolver de leur tempe.
MAKINE.A / LE TESTAMENT FRANÇAIS / 1995	Pp. 110-111	Ce n'est donc pas la variété des vins ou l'abondance rabelaisienne de la nourriture qui	comptaient	, mais... Nous pensions de nouveau à cet esprit français dont nous nous efforcions de percer le mystère.
MAKINE.A / LE TESTAMENT FRANÇAIS / 1995	Pp. 39-41	Seule la chronologie des longs récits de notre grand-mère	comptait	pour nous : un jour, dans leur temps légendaire, Paris surgissait des eaux, le soleil brillait et au même moment, nous entendions le cri encore lointain du train (…)
BOUDARD.A / MOURIR D'ENFANCE / 1995	Pp. 235-236	Elle devait se priver de quelque chose pour m'offrir le cinéma. C'est les seuls cadeaux qui	comptent	, ceux où l'on se prive pour offrir.
BOUDARD.A / MOURIR D'ENFANCE / 1995	Pp. 198-199	Ça me sert de support l'écriture... la main à la plume physiquement. Elle seule	compte	... le rythme... question d'oreille avant tout... qu'il n'y ait pas de couac...
BOUDARD.A / MOURIR D'ENFANCE / 1995	Pp. 180-181	Un billet... des pièces... j'allais me voler moi-même, il me prend en flag. Il appelle Bon Papa à la rescousse... ils	comptent	... mille huit cents balles... d'anciens francs ça fait pas lerche !
BOUDARD.A / MOURIR D'ENFANCE / 1995	Pp. 179-180	Elle surveillait tout, elle écoutait aux portes... elle	comptait	vos pas... vos économies... tout elle enregistrait, mieux que vos ordinateurs chéris.
BOUDARD.A / MOURIR D'ENFANCE / 1995	Pp. 166-167	Nous les copains, on ne	comptait	plus, tellement il était excité. Ce qui me revient à l'esprit pendant que ce prêtre ramollo nous convie au Pater noster.
BOUDARD.A / MOURIR D'ENFANCE / 1995	Pp. 25-26	J'ai aperçu le perron de loin, la façade couverte de lierre, ça ne m'a pas paru tout à fait le château de Chambord. Ce qui	compte	sans doute c'est l'image qu'on suscite. Monsieur d'Agrèves c'était le châtelain, le seigneur, une fois pour toutes.
BIANCIOTTI.H / LE PAS SI LENT DE L'AMOUR / 1995	Pp. 326-327	ils	comptaient	et recomptaient les sous que chacun avait glanés, avant de prendre leur invariable petit déjeuner de sardines à l'huile et de vin blanc.

Tableau 6 : Présentation des concordances (corpus littéraire)

Dans le tableau 7 nous présentons ces classes par effectif décroissant et nous y joignons un exemple issu du corpus. Le graphique (Figure 1) complète le tableau.

Construction	Fré-quence	Illustration
***N compte (pour N)* [importer]**	**38,0%**	(...) dans les rêves, à ce qu'il me semble, ces choses-là ne ***comptent*** pas. SALVAYRE.L /LA PUISSANCE DES MOUCHES/1995, Pages 93-94
***qn compte N<coll>* [dénombrer]**	**21,0%**	Rose Caterina claque-t-elle la portière, les enfants qui ***comptaient*** chacun leurs pièces de monnaie, se reprennent à quémander ; BIANCIOTTI.H /LE PAS SI LENT DE L'AMOUR/1995, Pages 16-17
qn compte INF [envisager de]	**9,0%**	J'espère, me dit A, que tu ne ***comptes*** pas laisser le rapport se rédiger tout seul ? ORMESSON.J D' /LA DOUANE DE MER/1993, Pages 187-188
qn compte sur qn [faire confiance à]	**8,0%**	elle la géra mal, se brouilla avec la famille, s'entoura enfin de jeunes gens au talent futur, qui ***comptaient*** sur elle pour vivre de leur art. Marie-Paule : subversion involontaire sur fond d'angoisse.KRISTEVA.J /LES SAMOURAIS/199,0 Pages 100-101
qc compte N<coll> [comprendre, avoir]	**8,0%**	Tout ce que son corps ***compte*** de noeuds, d'articulations et de cartilages, où s'amassent des flux comprimés, s'éveille maintenant et vibre. KRISTEVA.J /LES SAMOURAIS/1990, Pages 157-158
qn compte N (dans/parmi N) [inclure]	**3,5%**	C'est comme ça que j'ai appris qu'ils ***comptaient*** quelques Indiens parmi leurs lointains ancêtres et même un grand cacique. SEGUIN.F /L'ARME A GAUCHE/1990, Pages 131-132
qn compte sur qc [escompter qc]	**3,5%**	Le palais a besoin d'une plume. Je ***compte*** sur votre discrétion. Et sur une réponse demain neuf heures. ORSENNA.E /GRAND AMOUR/1993, Pages 50-52
qn compte [faire un décompte]	**3,0%**	Il y aura des étapes. Bien sûr, il y aura des étapes. Au bout de dix milliards d'années... - ça fait quelque chose, si je **compte** bien, comme cinq milliards d'années avant notre rencontre devant la Douane de mer ? dit A en baissant la tête sous l'effort. - Jour pour jour, lui dis-je. ORMESSON.J D' /LA DOUANE DE MER/1993 / Pages 156-157
qn compte N comme/pour [tenir N pour]	**2,0%**	Du reste, on y est dès le titre : Le Cousin Pons, k + z + s. Anagramme de Balzac, d'ailleurs, si l'on ***compte*** le c pour un k : Balzak-kas-noazet.... ' KRISTEVA.J /LES SAMOURAIS/1990, Pages 37-38
qn compte que P [escompter que]	**2,0%**	Micho Ludo ***compta*** qu'il était au Centre Saint-Paul depuis dix mois.QUEFFELEC.Y /LES NOCES BARBARES/1985, Page 236
N compte parmi N<coll> [être au nombre de]	**1,0%**	Les mois qui suivirent ce retour de vacances en Touraine compèrent parmi les plus exaltants que connut Olivier. SABATIER.R /LES FILLETTES CHANTANTES/1980/pp.208-212/
qn compte avec N [tenir compte de]	**1,0%**	L'éternel retour qui **compte** avec l'enfant et qui revient à lui pour le refaire est une idée d'éternité inconstante. KRISTEVA.J / LES SAMOURAIS / 1990 / Pages 258-259
à compter de N<date> [à partir de]	**0,0%**	

Tableau 7 : Fréquence des emplois du verbe ***compter*** dans le corpus littéraire

Il ressort des résultats de l'analyse du corpus que l'usage le plus fréquent du verbe ***compter*** dans le corpus littéraire est (avec 38%) l'expression d'un jugement d'importance ou de valeur attribué à une chose, un événement ou une personne par un juge qui n'est mentionné qu'occasionnellement. En l'absence du

complément *pour N*, le jugement est supposé valoir dans l'absolu. Dans le second exemple mentionné du tableau 7, le référent de « *ces choses-là* » est jugé sans importance par un juge indéterminé, mais relativement à un espace mental « *dans les rêves* ». Le second meilleur résultat (21%) est l'expression d'un acte de dénombrement. À partir de la troisième construction par fréquence décroissante, la fréquence tombe à moins de 10%.

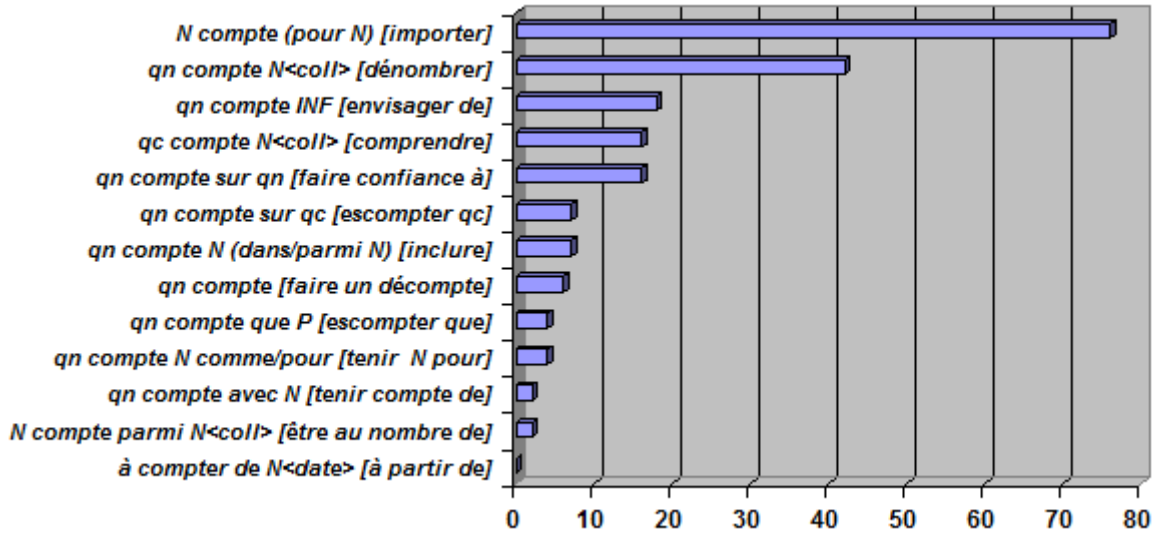


Figure 1 : Graphique de fréquence des emplois du verbe ***compter*** dans le corpus littéraire

1.4. Le profil des emplois du verbe *compter* dans le corpus journalistique

Dans le corpus journalistique (l'Est Républicain, mois de février 2002), on peut s'attendre à ce que la fréquence de certaines constructions du verbe compter diffère sensiblement. C'est effectivement le cas, comme en témoigne le tableau 8 des fréquences et le graphique (Figure 2) qui lui correspond :

Construction	Fré-quence	Illustrations
qc compte N<coll> [comprendre]	**25,0%**	Aujourd'hui, il y a 22.000 habitants et la mairie **compte** 600 agents.
à compter de N<date> [à partir de]	**22,0%**	Pour motif de congés, la mairie sera fermée **à compter de** ce mercredi 2 janvier jusqu'au samedi 5 janvier
qn compte N<coll> [dénombrer]	**15,0%**	La présidente elle-même ne **compte** pas le temps passé à préparer l'événement.
qn compte sur qc [escompter qc]	**10,5%**	Les salariés pourront **compter** sur au moins les trois quarts de leur rémunération.
qn compte INF [envisager de]	**9,0%**	Le gouvernement **compte** bien en 2002 lancer un important programme de réformes (…)
qn compte N (dans/parmi N) [inclure]	**6,5%**	« Nous, on travaille deux fois plus en **comptant** les réquisitions. Ce n'est pas normal »
N compte (pour N) [importer]	**5,0%**	Malheureusement ce second essai ne **comptait** pas pour le sprint du combiné nordique (…)
qn compte [faire un décompte]	**5,0%**	« Quand on aime gratter, on ne **compte** pas. Mon budget jeux, c'est ma folie à moi ».
qn compte sur qn [faire confiance à]	**1,0%**	(…) je sais **compter** sur vous pour accompagner ce projet (…)
N compte parmi N<coll> [être au nombre de]	**0,5%**	(…) Isabelle Delobel et Olivier Schoenfelder, « qui **comptent** parmi nos meilleurs amis... »
qn compte avec N [tenir compte de]	**0,5%**	(…) le Général Hiver nous prouve qu'il faut encore **compter** avec lui.
qn compte que P [escompter que]	**0,0%**	
qn compte N comme/pour [tenir N pour]	**0,0%**	

Tableau 8 : Fréquence des emplois du verbe ***compter*** dans le corpus journalistique

Cette fois-ci, c'est *qc compte N<coll>* [comprendre] qui figure en tête avec 25% des occurrences, suivi de près par *à compter de N<date>* [à partir de] avec 22% et de *qn compte N<coll>* [dénombrer] avec 15%. Les cinq constructions suivantes tournent autour de 5% à 10,5% et les cinq dernières autour de 0% à 1%.

L'emploi le plus fréquent réside donc dans l'établissement d'une simple relation entre un ensemble et une collection d'éléments (ex. *Aujourd'hui, il y a 22.000 habitants et la mairie compte 600 agents*.) L'article journalistique exploite donc prioritairement le verbe *compter* (en concurrence ici avec *avoir*, *comprendre* ou *comporter*, voire *enregistrer*) pour le dénombrement d'une collection de personnes, d'événements ou de choses.

Le second emploi est typique de la langue administrative pour exprimer le début d'un intervalle de temps, en général dans le futur (ex. *Pour motif de congés, la mairie sera fermée à compter de ce mercredi 2 janvier jusqu'au samedi 5 janvier*). Cela concerne les avis publics locaux, mais aussi les décisions administratives et politiques de tous niveaux.

Enfin la troisième construction par fréquence décroissante se distingue de la première par la sous-catégorisation [personne] du sujet (ex. « *Nous, on travaille deux fois plus en comptant les réquisitions. Ce n'est pas normal* »). La paraphrase n'est pas la même, comprendre (avoir, comporter) avec un sujet [chose] et inclure avec un sujet [personne].

Cependant des configurations intermédiaires se présentent. Ainsi la phrase *En France, on compte environ 36.000 communes* ne peut être paraphrasée ni par *comprendre / comporter*, ni par *inclure*, mais seulement par *avoir*. Le sujet est a priori un collectif de personnes, mais la phrase admet comme variante : *La France compte environ 36.000 communes, qui relèverait de la première construction.*

Il y a donc un glissement actanciel (non causatif → causatif) entre la simple relation partie-tout et l'action de dénombrement des parties du tout. Il est également à noter que la construction figée *sans compter N* occupe une place remarquable dans les 15% d'occurrences relevant de la construction plus générale *N<hum> compte N<coll>* avec 6 occurrences sur 30, soit 20%.

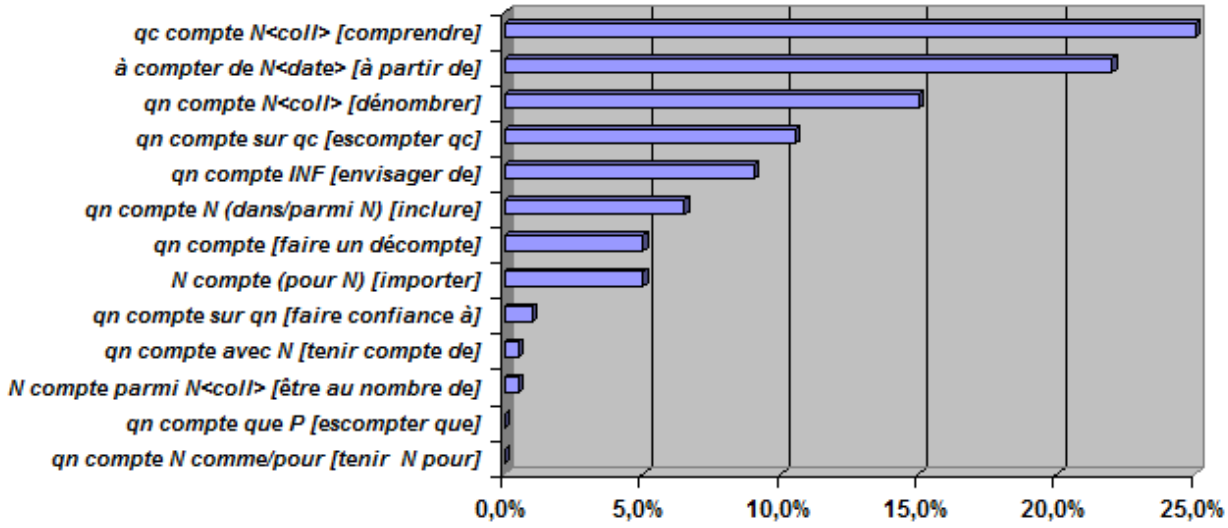

qc compte N<coll> [comprendre]
à compter de N<date> [à partir de]
qn compte N<coll> [dénombrer]
qn compte sur qc [escompter qc]
qn compte INF [envisager de]
qn compte N (dans/parmi N) [inclure]
qn compte [faire un décompte]
N compte (pour N) [importer]
qn compte sur qn [faire confiance à]
qn compte avec N [tenir compte de]
N compte parmi N<coll> [être au nombre de]
qn compte que P [escompter que]
qn compte N comme/pour [tenir N pour]
0,0%
5,0%
10,0%
15,0%
20,0%
25,0%

Figure 2 : Graphique de fréquence des emplois du verbe ***compter*** dans le corpus journalistique

1.5. Le profil des emplois du verbe *compter* dans le corpus politique

Notre corpus politique est composé de discours prononcé à la fin du 20e et au début du 21e siècle. La langue employée est semi-orale, le discours étant généralement écrit et souvent par un rédacteur patenté.

Construction	Fré-quence	Illustrations
N compte (pour N) [importer]	**36,5%**	Ce qui **compte** maintenant c'est de mettre en oeuvre cet accord. / Fillon 2010-01-05
qc compte N<coll> [comprendre]	**21,0%**	En 1999, il y a dix ans seulement, la Défense **comptait** 120 000 emplois. En 2007 160 000 / Dupon -2010-01-24
qn compte sur qn [faire confiance à]	**18,0%**	Nous avons besoin de vous pour sentir les vibrations de notre société. Je **compte** aussi sur vous pour ne pas laisser à la gauche le pouvoir de dire n'importe quoi / Fillon 2008-09-08
qn compte sur qc [escompter qc]	9,5%	Ce cadre est en train de se mettre en place. Je **compte** sur l'apport des états généraux pour en démultiplier l'efficacité / Fillon 2009-07-10
qn compte N<coll> [dénombrer]	7,0%	Les chiffres parlent d'eux - mêmes : nous **comptons** actuellement 4 retraités pour dix actifs. Chirac 2003-01-11
qn compte INF [envisager de]	4,0%	Sur ce sujet, comme sur tous les autres, je ne **compte** pas ce matin me contenter de pétitions de principe / Fillon 2004-10-28
qn compte que P [escompter que]	2,0%	Une partie de la société civile est lasse de l'isolement auquel ce régime la condamne. Nous **comptons** que les autorités iraniennes consentent enfin au dialogue / Fillon 2009-03-02
qn compte [faire un décompte]	1,0%	Si vous **comptez** bien, cela fait 50 000€ par emploi / Bayrou 2007-04-10
qn compte N (dans/parmi N) [inclure]	0,5%	Ce n'est pas à vous que je dois rappeler l'importance des activités liées à la mer. Aujourd'hui, en **comptant** le tourisme, ces activités emploient plus de 500 000 personnes / Fillon 2009-12-03
qn compte N comme/pour [tenir N pour]	0,5%	Je propose que, désormais, quand il s'agit de voter à l'Assemblée nationale, on ne puisse plus **être compté** comme votant si l'on est absent de l'hémicycle, que l'on ne puisse plus déléguer son vote (...) /Bayrou 2007-03-17
N compte parmi N<coll> [être au nombre de]	0,0%	
qn compte avec N [tenir compte de]	0,0%	
à compter de N<date> [à partir de]	0,0%	

Tableau 9 : Fréquences des emplois du Vb *compter* dans le corpus politique

La construction la plus fréquente dans le tableau 9 est la même que dans le discours littéraire avec une fréquence similaire : *N compte (pour N)* [importer] avec 36,5% (contre 38% dans le corpus littéraire), ce qui suggère que la notion d'importance (*N compte ≈ N importe / N entre en ligne de compte*) est prioritairement attachée au verbe compter dans ces deux types de discours, au détriment de ses emplois transitifs. En outre, on observe dans le corpus politique un trait du discours oral qui ne se retrouve dans le corpus journalistique qu'à travers des citations : il s'agit de la structure bipartite *ce qui / un N compte,// (c'est) SN<description>,* ex.

(4) Bien sûr cela **compte**, cela existe même pour les élèves des grandes écoles
(5) Au fond, il n'y a qu'une maxime qui **compte**, c'est "Qui n'est pas avec moi est contre moi".
(6) Pour moi, il n'y a qu'un sondage qui **compte**, c'est celui qui aura lieu les 22 avril et du 6 m
(7) Ce qui **compte**, c'est l'esprit d'équipe...
(8) ce n'est pas les individualités qui **compte** [sic], sans cela le Paris Saint-Germain serait champion (…)
(9) En effet, la seule chose qui **compte**, si nous voulons faire face, premièrement, (…) c'est (…)

La seconde construction par fréquence décroissante est, avec 21%, *qc compte N<coll>* [comprendre, avoir], c'est-à-dire la construction qui vient en tête dans le corpus journalistique (avec 25%). Cela suggère que le corpus politique présente pour les deux plus importantes constructions des traits qui le situent à mi-distance entre les deux autres corpus. Toutefois la troisième construction avec 18%, *qn compte sur qn* [faire confiance à] rapproche le corpus politique du littéraire, car dans celui-ci cette construction figure en 4e place avec 8% des occurrences, tandis que dans le corpus journalistique elle figure en queue de peloton avec seulement 1% des occurrences. Une caractéristique propre au corpus politique est un resserrement des constructions employées : plus de la moitié (7 sur 13) n'y figurent pas du tout ou atteignent au plus 2%.(cf. Figure 3).

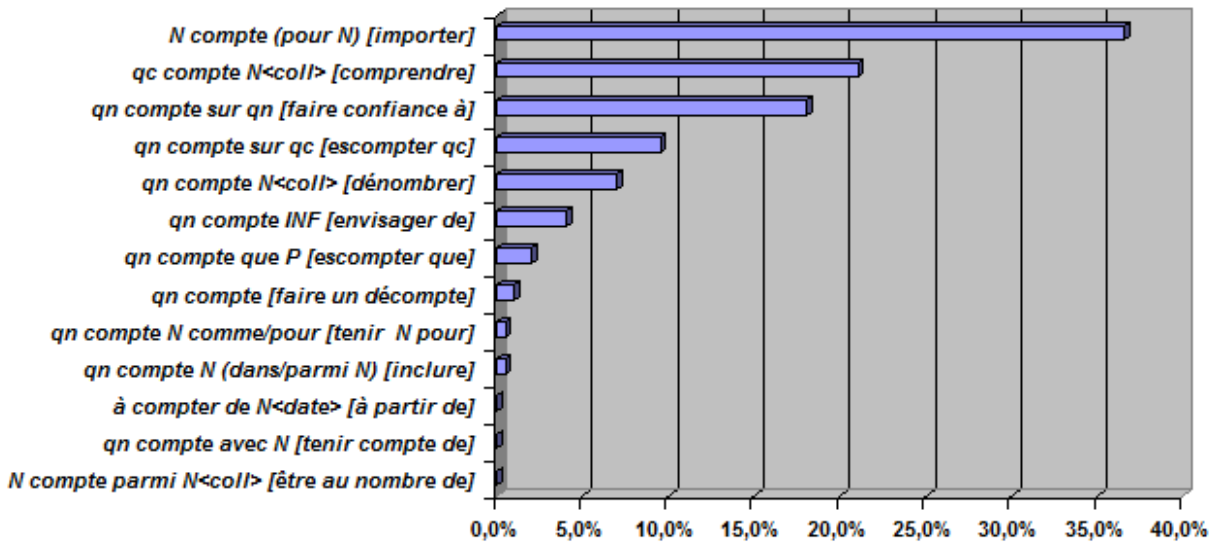


Figure 3 : Fréquence des emplois du Vb *compter* dans le corpus politique

1.6. Le profil comparé des emplois du verbe *compter* dans les trois corpus

Comparer le profil des emplois d'un verbe dans deux corpus est relativement aisé, il suffit de disposer les constructions selon leur différentiel de fréquence (cf. François 2007, Chapitre 4 ; François-Sandly 2007), ce qui fait immédiatement apparaître au moins les deux zones de priorités pour chaque corpus et un marais intermédiaire où les deux corpus se distinguent faiblement. La comparaison par cette méthode des deux corpus journalistique et politique fournit le résultat suivant (Tableau 10) :

	Est Républicain	**discours politiques**	**différentiel**	**zones**
***N compte (pour N)* [importer]**	**5,0%**	**36,5%**	**-31,5%**	**A**
***qn compte sur qn* [faire confiance à]**	**1,0%**	**18,0%**	**-17,0%**	
qn compte que P [escompter que]	0,0%	2,0%	**-2,0%**	**B**
qn compte N comme/pour [tenir N pour]	0,0%	0,5%	**-0,5%**	
N compte parmi N<coll> [être au nombre de]	0,5%	0,0%	**0,5%**	
qn compte avec N [tenir compte de]	0,5%	0,0%	**0,5%**	
qn compte sur qc [escompter qc]	10,5%	9,5%	**1,0%**	
qn compte [faire un décompte]	5,0%	1,0%	**4,0%**	
***qc compte N<coll>* [comprendre]**	**25,0%**	**21,0%**	**4,0%**	**C**
qn compte INF [envisager de]	9,0%	4,0%	**5,0%**	**B**
qn compte N (dans/parmi N) [inclure]	6,5%	0,5%	**6,0%**	
qn compte N<coll> [dénombrer]	15,0%	7,0%	**8,0%**	
***à compter de N<date>* [à partir de]**	**22,0%**	**0,0%**	**22,0%**	**D**

Tableau 10 : Comparaison par le différentiel de fréquence des fréquences des constructions du Vb *compter* entre les deux corpus journalistique et politique

Ce tableau comparatif révèle cinq zones : la zone A est celle des deux constructions fortement représentées dans le corpus politique et faiblement dans le corpus journalistique. Inversement la zone D est celle de la seule construction typique du corpus journalistique absente du corpus politique. Ces deux zones permettent de dégager les spécificités de chaque corpus. En revanche la partie centrale révèle, à côté de constructions faiblement représentées dans les deux corpus (zone B) une construction (zone C) bien représentée dans les deux corpus

et donc avec un différentiel faible. Cette zone C identifie la construction qui rapproche les deux corpus.

Dès qu'il s'agit cependant de comparer trois corpus, la méthode du différentiel de fréquence n'est plus exploitable. Le problème est alors de justifier l'ordre dans lequel les différentes constructions vont être examinées.

La solution que nous proposons ici consiste à faire comme si nous avions étudié les 600 occurrences des trois corpus de manière indistincte, de manière à dégager les constructions qui, globalement, sont bien représentées, et celles qui ont un score faible, quel que soit le corpus. Pour cela il faut calculer la moyenne des fréquences des trois corpus et disposer les constructions par fréquence moyenne décroissante. C'est ce qui est fait dans le tableau 11.

	FRANTEXT	**Est Républicain**	**Discours politiques**	**MOYENNE**	
***N compte (pour N)* [importer]**	**38,0%**	**5,0%**	**36,5%**	**26,5%**	**Excellence FRANTEXT & DisPol**
***qc compte N<coll>* [comprendre]**	**8,0%**	**25,0%**	**21,0%**	**18,0%**	**Priorité EstRép & DisPol**
***qn compte N<coll>* [dénombrer]**	**21,0%**	**15,0%**	**7,0%**	**14,3%**	**Priorité FRANTEXT& EstRép**
qn compte sur qn [faire confiance à]	8,0%	1,0%	**18,0%**	9,0%	Priorité DisPol
qn compte sur qc [escompter qc]	3,5%	10,5%	9,5%	7,8%	Priorité EstRép & DisPol
à compter de N<date> [à partir de]	0,0%	**22,0%**	0,0%	7,3%	Priorité EstRép
qn compte INF [envisager de]	9,0%	9,0%	4,0%	7,3%	
qn compte N (dans/parmi N) [inclure]	3,5%	6,5%	0,5%	3,5%	
qn compte [faire un décompte]	3,0%	5,0%	1,0%	3,0%	
qn compte que P [escompter que]	2,0%	0,0%	2,0%	1,3%	
qn compte N comme/pour [tenir N pour]	2,0%	0,0%	0,5%	0,8%	Faibles scores partagés
N compte parmi N<coll> [être au nombre de]	1,0%	0,5%	0,0%	0,5%	
qn compte avec N [tenir compte de]	1,0%	0,5%	0,0%	0,5%	

Tableau 11 : Fréquence comparée des constructions du Vb *compter* dans les trois corpus

Ce tableau révèle une première zone d'excellence pour le corpus littéraire et le corpus politique pour la construction N compte (pour N), laquelle est en revanche peu représentée dans le corpus journalistique. Ensuite on rencontre une zone de priorité pour les deux corpus journalistique et politique avec la construction *qc compte N<coll>* [comprendre, avoir] et une zone de priorité pour les deux corpus littéraire et journalistique avec la construction *qn compte N<coll>* [dénombrer]. Dans aucun des trois cas saillants, une construction n'arrive à s'imposer sur la base de sa fréquence dans un seul corpus et ces trois constructions représentent les zones de priorité conjointe dans les corpus pris deux à deux. Dans le groupe des trois zones qui suivent, le corpus littéraire

ayant déjà livré 67% de ses occurrences, les trois constructions constituent une priorité du corpus politique (*qn compte sur qn* [faire confiance à]), du corpus journalistique (*à compter de N<date>* [à partir de]) ou des deux à la fois (*qn compte que qc* [escompter qc]) en présence d'une fréquence faible dans le discours littéraire (3,5%). Le troisième groupe de constructions, fort de 7 constructions, rassemble celles dont le score est faible dans les corpus (moins de 10%). À ce titre, ces constructions peuvent être considérées comme marginales.

Les données du tableau 11 se retrouvent dans la figure 4 sous la forme d'un graphique constitué de trois courbes et combiné à la table des données. Pour une meilleure lecture nous disposons le graphique en 'portrait' :

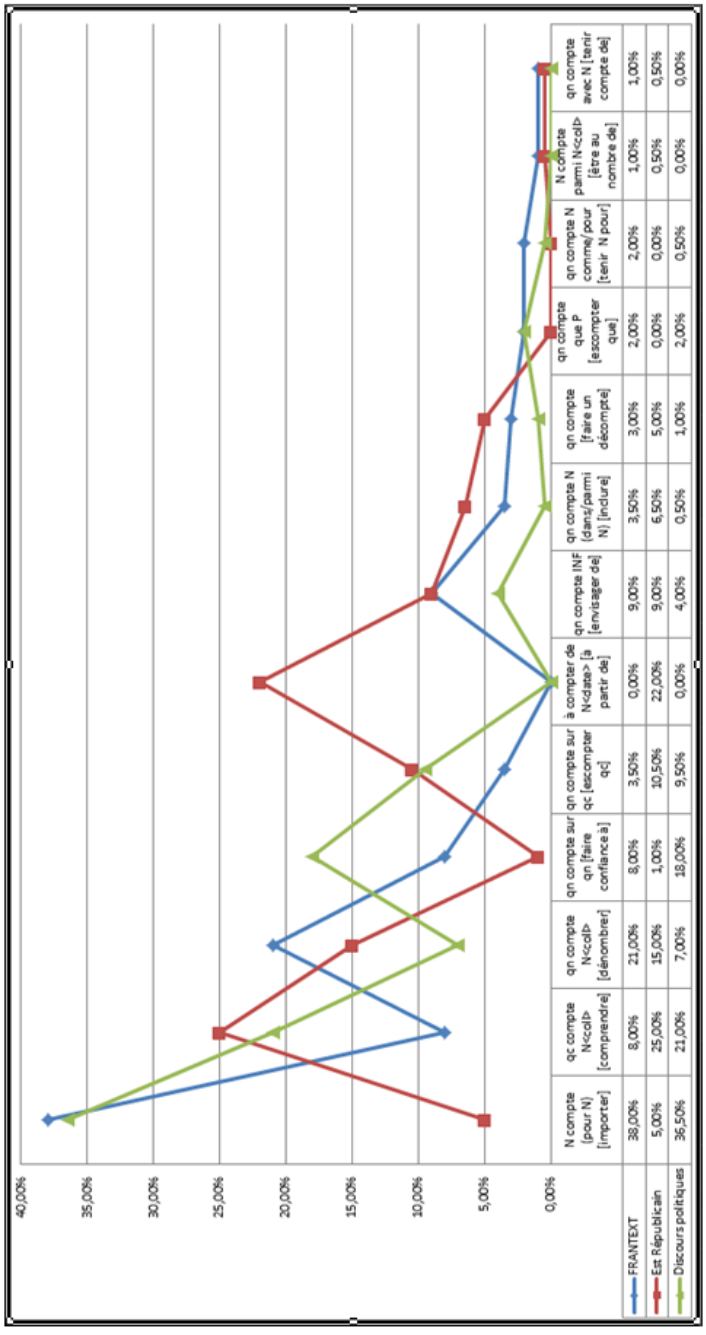

40,00%
35,00%
30,00%
25,00%
20,00%
15,00%
10,00%
5,00%
0,00%
FRANTEXT
Est Républicain
Discours politiques
N compte (pour N) [importer]
qc compte N<coll> [comprendre]
qn compte N<coll> [dénombrer]
qn compte sur qn [faire confiance à]
qn compte sur qc [escompter qc]
à compter de N<date> [à partir de]
qn compte INF [envisager de]
qn compte N [dans/parmi N] [inclure]
qn compte [faire un décompte]
qn compte que P [escompter que]
qn compte N comme/pour [tenir N pour]
N compte parmi N<coll> [être au nombre de]
qn compte avec N [tenir compte de]
38,00% 8,00% 21,00% 8,00% 3,50% 0,00% 9,00% 3,50% 3,00% 2,00% 2,00% 1,00% 1,00%
5,00% 25,00% 15,00% 1,00% 10,50% 22,00% 9,00% 6,50% 5,00% 0,00% 0,00% 0,50% 0,50%
36,50% 21,00% 7,00% 18,00% 9,50% 0,00% 4,00% 0,50% 1,00% 2,00% 0,50% 0,00% 0,00%

Figure 4 : Fréquences comparées des emplois du Vb *compter* dans les trois corpus

1.7. Conclusion intermédiaire

On l'aura sans doute compris, notre propos, avec cette analyse comparative de trois corpus, est de plaider contre une analyse de corpus indifférenciée, dès lors qu'elle a une ambition quantitative. Il suffit pour s'en convaincre d'examiner dans le tableau précédent les scores des deux constructions qui sont caractéristiques d'un corpus particulier (cf. Tableau 12) :

	FRANTEXT	**Est Républicain**	**Discours politiques**	**MOYENNE**
qn compte sur qn [faire confiance à]	8,0%	1,0%	**18,0%**	**9,0%**
à compter de N<date> [à partir de]	0,0%	**22,0%**	0,0%	**7,3%**

Tableau 12 : Illustration de la nécessité d'une analyse comparative de corpus

Quel peut être l'intérêt du score moyen de 9,0% pour *qn compte sur qn* [faire confiance à] et de 7,3% pour *à compter de N<date>* [à partir de] ? Ces scores ne sont pas interprétables en soi, car ils résultent du nivellement de scores significatifs dans l'un des trois corpus face aux deux autres. Dans le premier cas, la construction se révèle typique du corpus politique, peu représentée dans le corpus littéraire et négligeable dans le journalistique. Dans le second, c'est encore plus manifeste, comme nous venons de le voir sur le graphique précédent, puisque la construction figée *à compter de N<date>* représente plus d'un cinquième des occurrences du corpus journalistique et ne figure dans aucun des deux autres.

On notera par ailleurs que nous n'avons nulle part évoqué une validation **STATISTIQUE**. Une analyse quantitative est pertinente pour des corpus de taille relativement modeste comme ceux que nous avons examinés ici avec seulement 200 occurrences par corpus, toutes classées à la main (ce qui s'est révélé indispensable dans certains cas où la sélection des actants ne suffit pas à déterminer une classe de constructions et où l'ensemble de la signification de la phrase doit être imaginé et paraphrasé pour lever les doutes).

Une analyse statistique calcule, par différents procédés mathématiques, la fiabilité des moyennes des valeurs attribuées aux éléments d'un ensemble supposé homogène. Considérons les scores des trois corpus représentés dans le tableau 11. Quelles moyennes (colonne de droite) sont-elles fiables ? Un calcul statistique montrera très certainement que seules les constructions figurant dans les quatre dernières lignes qui ont un score moyen entre les trois corpus compris

entre 0,5% et 1,3% et dont aucun score dans l'un ou l'autre des corpus ne dépasse 2% ont un score moyen 'significatif', c'est-à-dire fiable (sur l'exploitation du test du χ2, on pourra consulter le chapitre 4 du Cahier 29 du CRISCO).

Si nous étendions cette étude à des corpus plus conséquents, de l'ordre de 1000 occurrences par corpus, il faudrait peut-être alors rechercher des sous-classes de corpus susceptibles de fournir des scores significativement distinctifs, par exemple pour le corpus journalistique en fonction de la rubrique (politique, faits divers, sports, économie, etc., cf. François 2010b) ou pour le corpus politique en fonction de la distinction gauche, centre et droite. Mais au niveau où nous nous sommes positionnés, cette sous-classification ne parait pas encore nécessaire.

II. TROIS OUTILS ÉLÉMENTAIRES DE LINGUISTIQUE QUANTITATIVE

L'objet de cette seconde partie est de fournir d'abord quelques ressources textuelles et, ensuite, de décrire en détail notre méthode d'exploitation de trois outils :

a) le **concordancier** AntConc,

b) deux outils de Microsoft Office :

- le **logiciel de traitement de texte** WORD (éventuellement remplaçable par la composante Writer de la plateforme Open Office téléchargeable gratuitement)

- le **tableur** EXCEL(éventuellement remplaçable par la composante Calcul de la même plateforme)

Nous souhaitons ainsi permettre au lecteur intéressé de reproduire les résultats présentés en première partie et de procéder à d'autres analyses comparatives de corpus selon une procédure similaire. Le contenu des nombreuses saisies d'écran étant immédiatement explicité par le contexte, ces figures ne sont pas numérotées.

II-1. LES BASES TEXTUELLES

Le linguiste qui cherche à observer la langue et à étudier les contextes d'emploi de mots ou d'expressions sur un grand nombre de textes trouve sur la Toile plusieurs bases textuelles numérisées (bibliothèques virtuelles), et plus ou moins prêtes au dépouillement. Ces textes sont très variés. On en trouve de tous les genres : des textes littéraires, des discours politiques, des transcriptions de l'oral, des articles de journaux, etc.

Certains sites proposent de fournir au chercheur des concordances en lignes, soit à partir de leurs propres bases de textes et de leur concordancier, soit depuis les fichiers que l'utilisateur possède sur son ordinateur.

II-1.1. Sites de recherche en ligne

Nous avons sélectionné quelques-uns de ces sites où le chercheur, qui ne possède pas de textes sur son propre disque, peut utiliser les bases textuelles littéraires ou journalistiques mises à sa disposition par le site.

A. Sur des corpus littéraires

a) FRANTEXT

La base est constituée de textes français qui s'échelonnent du XVIe au XXe siècle et est accessible sur abonnement. Certaines parties sont mises à la disposition des étudiants de l'agrégation.

Adresse ► http://www.atilf.fr

b) Rabelais et son temps

Le vocabulaire de Rabelais, son évolution, sa structure et les distances lexicales des mots sont consultables à l'adresse :

► http://134.59.31.3/rabelais.html

c) La Comédie humaine de Balzac

Le site offre une concordance en ligne, limitée aux 360 premières occurrences de chaque forme à l'adresse :

► http://134.59.31.3/~brunet/BALZAC/BALZAC.htm

d) Correspondance de Flaubert

Toute la correspondance de Flaubert y est.

Adresse ► http://www.univ-rouen.fr/flaubert/

e) Jean de La Fontaine

Le site offre l'œuvre complète de La Fontaine (fables, contes, théâtre, poèmes, prose) et un appareil critique. L'utilisateur a la possibilité de télécharger les Fables et un logiciel de lecture proposé par le site (WinFables). Le site comprend un moteur de recherche interne.

Adresse ► http://www.lafontaine.net/

f) Guy de Maupassant

Un moteur de recherche fonctionne sur l'ensemble de l'œuvre de Guy de Maupassant à l'adresse ► http://maupassant.free.fr

B. Sur des corpus journalistiques

a) GlossaNet en ligne

Le logiciel GlossaNet permet après enregistrement gratuit sur le site, de choisir parmi 80 journaux quotidiens en ligne, édités en 5 langues, d'enregistrer une requête qui sera appliquée quotidiennement sur l'édition du jour. Les résultats de ces requêtes sont envoyés quotidiennement par courrier électronique sous forme de concordances à l'utilisateur.

Adresse ► http://glossa.ladl.jussieu.fr.

b) Le journal *Le Monde*

Le journal "Le Monde" offre un outil linguistique gratuit qui peut être utilisé, selon ses concepteurs, à la fois comme un vérificateur d'orthographe, une banque lexicographique des collocations, ou encore comme un recueil des paradigmes morphologiques et de leurs réalisations contextuelles.

Adresse ► http://www.bultreebank.org/french/login.html

c) Concordancier-corpus français

Adresse ► http://www.lextutor.ca/concordancers/concord_f.html

C. Sur des corpus politiques

L'excellent site de Jean Véronis offre un moteur de recherche sur presque tous les discours politique français et notamment ceux des deux dernières campagnes présidentielles.

► http://www.veronis.fr/outils

II-1.2. Sites de téléchargement de textes numérisés

Si le chercheur préfère mener ses propres recherches sur des bases autres que celles qui lui sont imposées par les plateformes, il lui est possible de télécharger des milliers de textes à partir de bases textuelles en téléchargement gratuit et de les manipuler avec des concordanciers également gratuits.

A. Les bases textuelles CNRTL

Le Centre National de Ressources Textuelles et Lexicales (CNRTL) permet l'accès à plusieurs corpus, Adresse : http://www.cnrtl.fr/corpus/ entre autres au :

a) Corpus journalistique de l'Est Républicain

Ce corpus est constitué des données textuelles correspondant à trois années (1999, 2002 et 2003) de toutes les éditions intégrales du quotidien régional.

b) Corpus Oraux en Français (TCOF)

La base mise à la disposition de la communauté scientifique comprend les corpus oraux constitués dans les années 80-90.

c) Corpus d'articles de linguistiques issus de la revue "Sciences Humaines"

La revue "Sciences Humaines" diffuse ces articles sous la licence créative Commons (attribution à l'auteur, pas d'utilisation commerciale, rediffusion aux mêmes conditions).

d) DEDE : un corpus annoté pour le traitement des DEscriptions DEfinies

Le corpus DEDE est composé d'une série d'articles du journal *Le Monde* datant de septembre 1987 et appartenant à toutes les rubriques.

B. Corpus oraux

a) CFPP2000, composé d'un ensemble d'interviews sur les quartiers de Paris et de la proche banlieue, accessibles sans restriction.

Adresse ► http://ed268.univ-paris3.fr/CFPP2000/.

b) Corpus de français parlé

Elicop (Etude LInguistique de la COmmunication Parlée) regroupe les projets Elilap (Etude LInguistique de la LAngue Parlée) et Lancom (LANgue et COMmunication). C'est la transcription d' un grand corpus informatisé de la langue parlée.

Adresse ► http://bach.arts.kuleuven.be/elicop/

C. Bibliothèques virtuelles

a) Gallica

Les ressources de la Bibliothèque nationale de France (BnF) sont accessibles sur le serveur Gallica.

Adresse ► http://gallica.bnf.fr/

b) Le Net des Etudes françaises

La base propose des textes, entretiens, outils d'analyse, galerie,... et offre des liens notamment sur une base recensant le théâtre en ligne (1600-1800).

Adresse ► http://www.etudes-francaises.net/

c) ABU (Association des Bibliophiles universels)

La base, un corpus de textes francophones, hébergée par le CNAM, comprend en janvier 2002 : 288 textes de 101 auteurs.

Adresse ► http://abu.cnam.fr/

d) ATHENA

Ce site de l'Université de Genève recense les éditions électroniques des écrivains suisses de langue française, notamment Madame de Charrière, Jean-Jacques Rousseau, Benjamin Constant, Madame de Staël, etc.

Adresse ► http://un2sg4.unige.ch/athena/html/swis_txt.html

e) Poésie française

Une anthologie de la poésie française (5900 poèmes), mise en ligne par Webnet.

Adresse ► http://poesie.webnet.fr/

II-2. CONCORDANCIER ET ÉDITEURS DE TEXTES

Sur Internet, les concordanciers sont désormais très nombreux. Il suffit de saisir dans Google le mot « concordancier » pour voir apparaitre pas moins de cinq concordanciers sur la seule première page.

Pour notre démonstration, nous avons opté pour le concordancier gratuit **AntConc** et pour les éditeurs de textes WORD et EXCEL de Microsoft Office[12].

Nous avons essayé de suivre les procédures d'utilisation des logiciels et à chaque étape nous avons réalisé des captures d'écrans sur lesquelles nous avons noté l'ordre des étapes de la manipulation.

II-2.1. Le concordancier AntConc

AntConc est un logiciel de type concordancier, développé par le Professeur Laurence Anthony, de la Faculté des Sciences et de génie, Université de Waseda, Japan. Il est téléchargeable gratuitement sur le site de son auteur[13] et tourne sous Windows, MacOS X et GNU/Linux. Il ne nécessite pas d'installation et se lance en double cliquant simplement sur le fichier exécutable (.exe). Sa dernière version **AntConc3.3.1w** (beta) est encore à l'essai, mais fonctionne correctement avec *Windows XP, Vista* et *Win 7.*

Ce concordancier ne permet qu'une recherche morphologique, c'est-à-dire que le programme ne recherche dans les textes numérisés que des formes alphanumériques[14]. Juste la forme saisie fait l'objet d'une recherche, la moindre faute de frappe dans le texte numérisé ou dans la requête peut générer un résultat nul ou erroné.

---

[12] Pour éviter les problèmes d'incompatibilité entre les différentes versions de WORD et EXCEL, Microsoft Office, sur son site, propose deux solutions :
1) si le document est créé (nous utilisons, pour notre démonstration, la version MSOffice 2010) à destination de personnes qui utilisent des versions antérieures à celle de votre WORD, il faudra vérifier si elles ont installé le module de compatibilité Microsoft Office pour les formats de fichiers WORD, EXCEL et PowerPoint Open XML, dans ce cas le travail en mode Microsoft WORD 2010 ne pose aucun problème.
2) Si ces personnes n'ont pas installé le module de compatibilité Microsoft Office pour les formats de fichiers WORD, EXCEL et PowerPoint Open XML, le travail doit se faire en mode de compatibilité. Ce mode de compatibilité garantit qu'aucune fonctionnalité, nouvelle ou améliorée, de WORD 2010 n'est disponible lorsque vous travaillez dans un document. De cette façon, les personnes qui utilisent des versions antérieures de WORD bénéficieront de toutes les fonctionnalités d'édition.
Pour plus d'information, consulter les pages suivantes :
http://www.microsoft.com/downloads/fr-fr/details.aspx?familyid=941b3470-3ae9-4aee-8f43-c6bb74cd1466
http://www.commentcamarche.net/download/telecharger-34055375-pack-de-compatibilite-microsoft-office

[13] http://www.antlab.sci.waseda.ac.jp/software.html

[14] Il existe des programmes, tel Unitex, qui permettent des recherches catégorisées

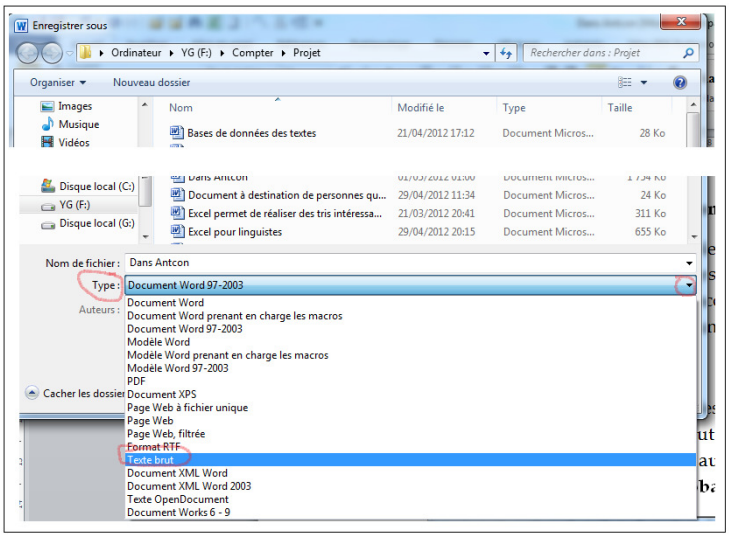
Pour optimiser la recherche avec **AntConc**, il est important de connaître les options de configuration que le logiciel offre à l'utilisateur.

A. Configuration de AntConc

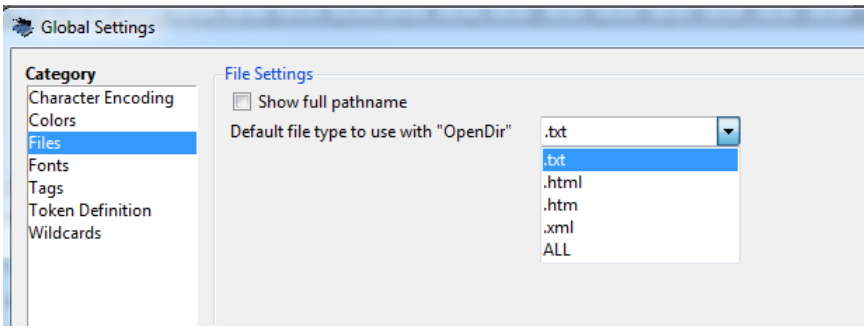
AntConc comporte des réglages préalables qui permettent de préparer en amont le corpus des concordances. Ces réglages concernent 1) les formats de fichier compatibles avec le programme, 2) l'encodage du texte, et 3) la préparation des énoncés du corpus à prendre la forme d'un tableau.

a) Les formats de fichier

AntConc fonctionne, par défaut, sur des fichiers au format « Texte brut », reconnaissable à leur extension (.txt)[15]. Pour obtenir un fichier (.txt) à partir de WORD de Microsoft Office, il suffit de l'**enregistrer sous** le même nom, et choisir le **type** *Texte brut*. Un nouveau fichier est créé avec l'extension .txt



Mais le programme peut être réglé pour accepter d'autres formats de fichiers : .htm, .html, .xml. L'option ALL autorise le travail sur tous les types de fichiers, entre autres les fichiers .doc de WORD. Le réglage est accessible via le menu **Global Setting** > **Category** > **Files**.

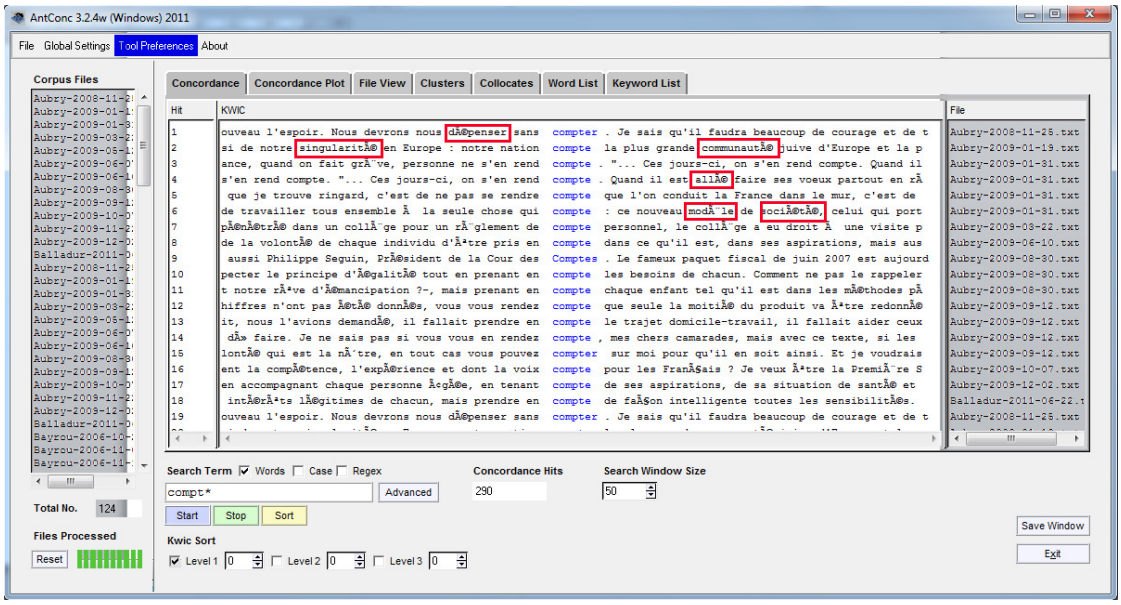


b) L'encodage du texte

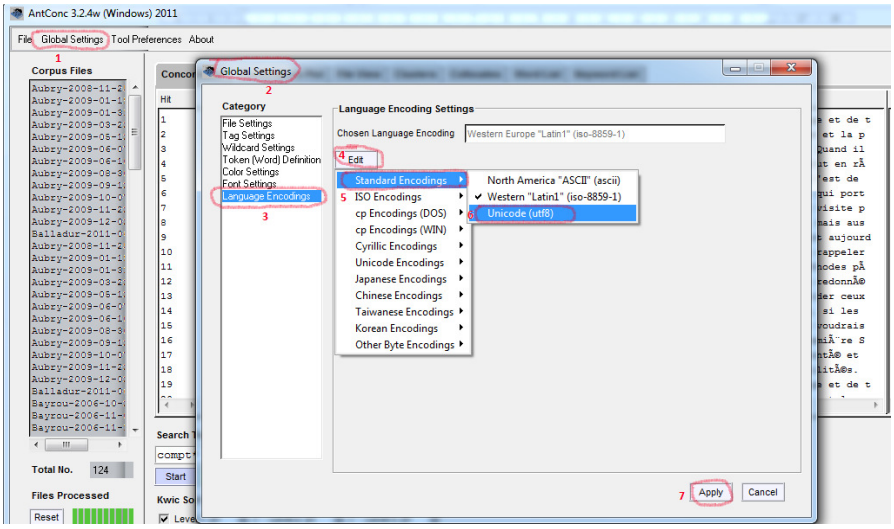
Parfois, la lecture du code peut être grandement pénalisée si les fichiers sont édités avec des encodages différents. En effet les voyelles accentuées et autres caractères particuliers, comme 'ç', apparaissent sous la forme de carrés ou autres bizarreries. Ces différences d'encodage peuvent être pénalisantes quand

[15] L'éditeur de textes Notepad++, téléchargeable gratuitement à l'adresse http://notepad-plus-plus.org/ est un excellent concurrent au très rudimentaire Bloc-notes de Windows.

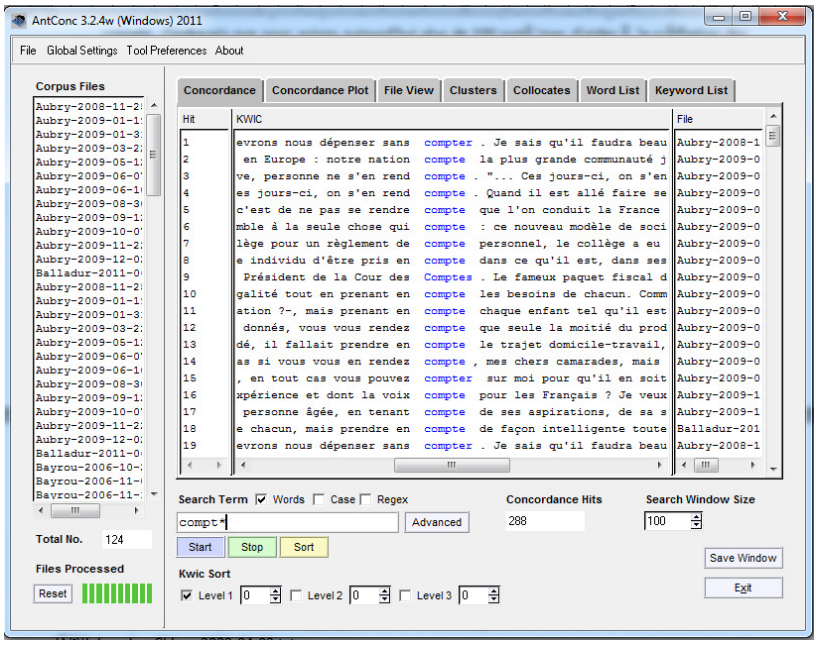
on utilise des fichiers mal convertis avec certains logiciels ou lors des téléchargements à partir de sites Internet



Ce problème est prévu par le concepteur d'**AntConc**. Pour le corriger, il faut, avant de lancer une requête, configurer le logiciel. Dans le menu **Global Setting** (1) aller à l'option **Language Encodings** (3) puis avec le bouton **Edit** (4) ouvrir **Standard Encodings** (5) et choisir **Unicode** (utf8) (6), enfin valider (7) pour appliquer la nouvelle configuration.

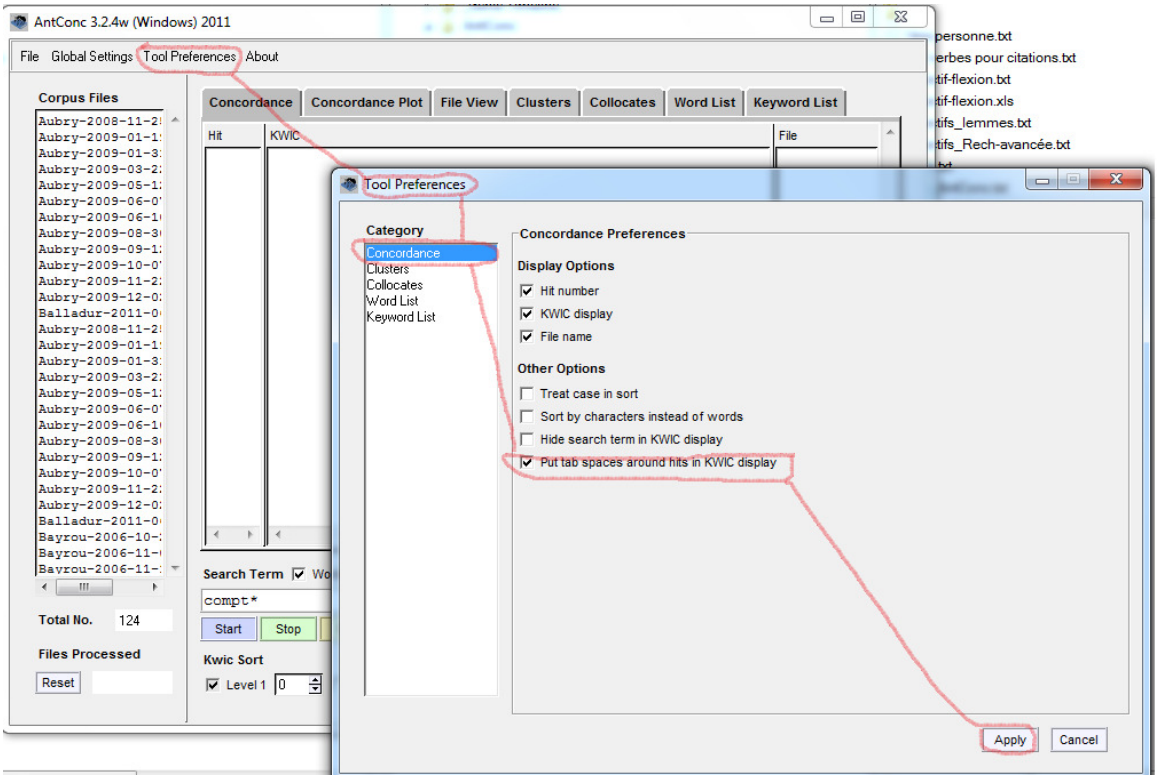


Une fois le bon encodage appliqué, le résultat présente un corpus lisible et prêt à l'utilisation.



c) La préparation des énoncés du corpus

L'option **Put tab spaces around hits…,** accessible depuis le menu **Tool Preferences** > **Concordances**, place une tabulation avant et une autre après l'unité ou la suite d'unités recherchées. Ces tabulations permettent, ultérieurement, de construire automatiquement dans WORD ou EXCEL des tableaux, afin de profiter des différentes possibilités de tris offertes par ces deux éditeurs.



En effet, à leur sortie, les énoncés obtenus comportent de part et d'autre des unités recherchées un espace qui sera converti en tabulation, dans WORD**,**

Notepad++, ou tout autre éditeur. Ces tabulations correspondent aux lignes de séparation des colonnes dans un tableau. Une simple addition (*n* tabulations + 1) donne le nombre de colonnes qu'on obtiendrait dans un tableur, comme EXCEL, par exemple.

En outre, par défaut, **AntConc** numérote les énoncés et place à la fin de chaque énoncé le nom du fichier d'où il est extrait.

B. Les requêtes

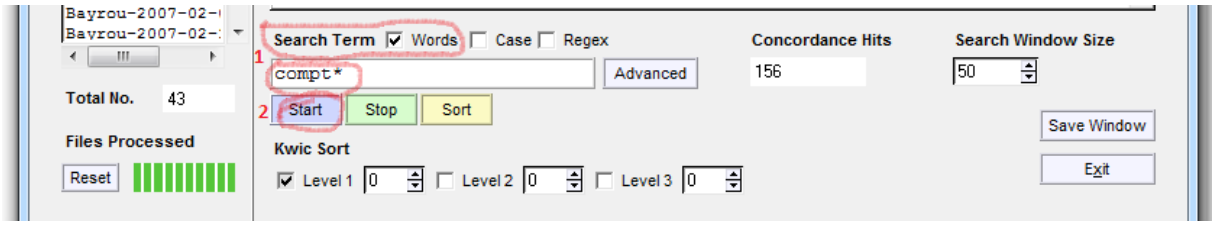
Dans la fenêtre principale, plusieurs options s'offrent.

Avec l'option **Search Window side**, il est possible de contrôler le nombre de caractères de texte devant être relevés de chaque côté du terme recherché. La valeur de défaut est 50 caractères.

La génération des concordances peut être interrompue à tout moment en cliquetant sur le bouton **Stop**.

a) Requête simple

La recherche des termes se fait dans un fichier unique ou dans un dossier contenant plusieurs fichiers. La requête est saisie directement au clavier dans la fenêtre principale du programme en dessous de **Search Term**, ou copiée d'un autre emplacement, par copier-coller.



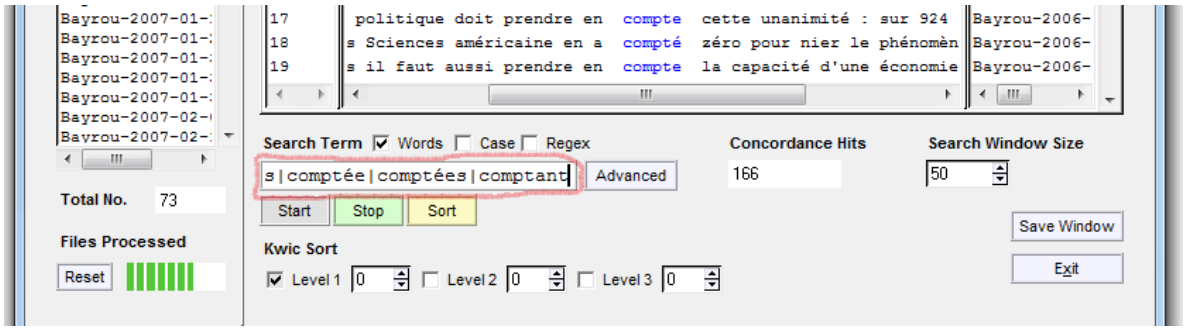
Avec les verbes réguliers du premier groupe, par exemple, la requête avec le joker (*) est simplifiée par le fait que le verbe présente la même base dans toute sa flexion. Il suffit donc d'inscrire la base suivie de l'astérisque (**compt***, dans notre cas) pour obtenir tous les énoncés qui comportent un mot commençant par COMPT. Cependant les résultats d'une requête de ce type (joker [.*]) génère très souvent des énoncés non pertinents. Dans notre exemple, nous avons demandé tous les énoncés qui comportent la base du verbe *compter* en inscrivant, dans la fenêtre des requêtes, **compt***. Le corpus comprend des énoncés dont des mots-clés comme *comptable, comptabiliser*, *comptabilité*, etc., sont, pour notre recherche, considérés comme des mots "parasites". L'utilisateur saura écarter ultérieurement ces résultats non pertinents avec le tri alphabétique dans EXCEL.

b) Requête multiple

Avec **Antconc,** on peut éviter ces mots "parasites", en inscrivant, directement dans le champ des requêtes (**Search Term**) de la fenêtre principale, tout le paradigme de conjugaison d'un verbe pour empêcher le programme de relever les mots qui commencent de la même manière que la base du verbe. La

saisie de la liste des formes du verbe *compter* par exemple se fait comme suit : on inscrit les différentes formes du verbe séparées par un slash droit (|)[16] :

> compter|comptaient|comptais|comptait|comptai|comptâmes|comptassent|comptasses|comptasse|comptassiez|comptassions|comptas|comptâtes|comptât|compta|comptent|comptent|compteraient|compterait|compterai|compteras|comptera|comptèrent|compterez|compteriez|compterions|compterons|compteront|comptes|compte|comptez|comptiez|comptions|comptons|compterais|compté|comptés|comptée|comptées|comptant



On peut trouver sur la toile des sites[17] qui offrent des listes de mots très utiles pour les requêtes, pour les comparaisons ou les statistiques. Le site d'un blog à l'adresse (http://blog.bafouillages.net/2009/11/analyse-lexicale-d-un-texte-ou-d-un-discours-avec-antconc/) donne pour AntConc quatre fichiers en Texte brut qui contiennent les mots français avec leurs différentes flexions : un fichier pour les noms (Nlem.txt), un deuxième pour les verbes (Vlem.txt) et un troisième pour les adjectifs (Alem.txt) ; le quatrième et dernier fichier rassemble les trois premiers en un seul document (Dico_lemmes)[18].

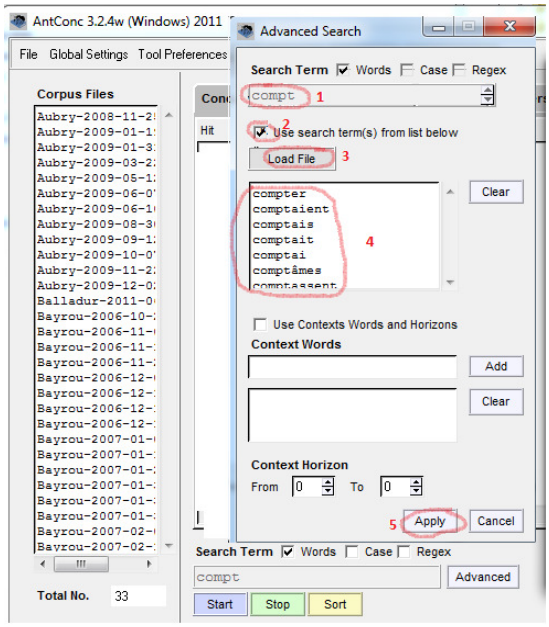
c) Recherche avancée

Une autre façon pour lancer une requête multiple est possible avec le menu **Advanced**, accessible à partir de la fenêtre principale. Quand on appelle cette procédure, une fenêtre de dialogue **Advanced Search** apparait. Dans la rubrique **Search Term**, On inscrit un mot quelconque (1), de préférence en rapport avec la requête, par exemple dans notre cas, COMPT, puis (2) on coche l'option **Use search term(s) from list below**. Deux possibilités sont offertes : ou bien on saisit les mots, un par la ligne, ou bien on charge une liste de termes en indiquant au programme, par un clic sur le bouton **Load File** (3) (ouvrir fichier), le chemin d'accès d'un fichier (.txt). Une seconde fenêtre, propre à Windows, guide vers le dossier puis vers le fichier de format .txt, qui aura été créé au préalable et qui contiendrait, sous forme de liste (un mot par ligne), le paradigme de mots à rechercher. Dans notre cas, il s'agit des toutes les formes du verbe *compter*.[19]

---

[16] Le slash droit s'obtient avec la combinaison des deux touches AltGr + 6.
[17] http://eduscol.education.fr/cid50486/liste-de-frequence-lexicale.html ; http://www.lexique.org/ ; http://www.cnrtl.fr/lexiques/morphalou/
[18] Ce document est utile pour une recherche lemmatisée avec AntConc.
[19] Pour obtenir cette liste, WORD permet de remplacer le slash droit par un retour à la ligne en remplaçant le slash (|) par le symbole (^p). L'opération de remplacement fonctionne dans les deux sens..

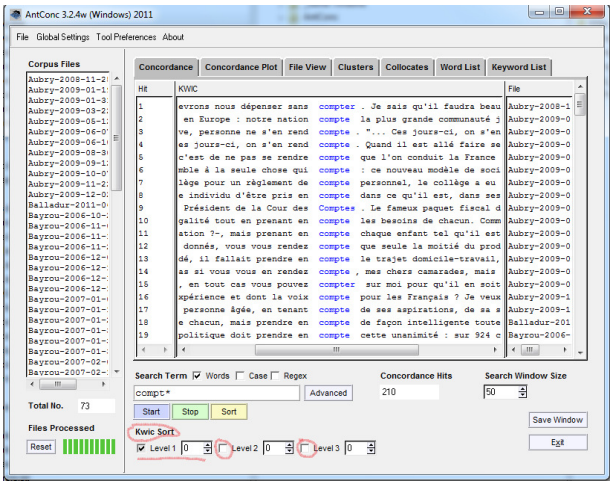
Une fois le choix fait, sur la fenêtre de dialogue **Advanced Search** apparait le début de la liste en question (4). On valide (5) l'option en appuyant sur **Apply** pour retourner à la fenêtre d'**AntConc**. Le nom qu'on a donné à la liste (**compt**) apparaît alors dans la fenêtre des requêtes désormais inutilisable. La recherche peut être lancée enfin en cliquant sur **Start**.



C. Les tris

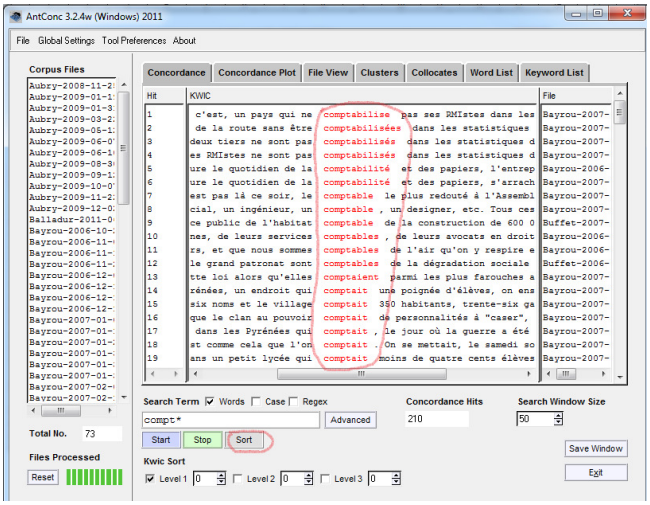
AntConc permet entre autres de gagner du temps en réalisant des tris avant d'enregistrer le fichier des résultats.

AntConc permet de trier (Kwic Sort)[20] les résultats sur trois niveaux (Level 1, Level 2 et Level 3) réglables à souhait. Chaque niveau permet un tri, visible sur la fenêtre principale grâce à une couleur différente.

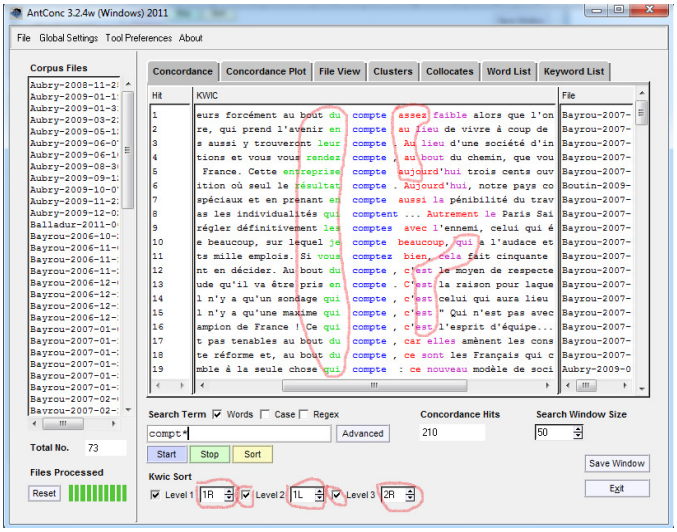


[20] **Kwic Sort** "sort by Key word in context", qu'on peut traduire par : « tri par mot clef en contexte ».

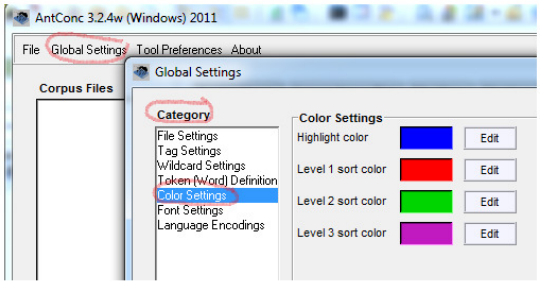
Au niveau 1, la valeur 0 opère sur la liste des mots recherchés un classement par ordre alphabétique, **1R** (R pour l'anglais Right = droite) trie le corpus selon le premier mot immédiatement placé à droite du mot recherché, **1L** (L pour l'anglais Left = gauche) classe les énoncés selon le premier mot immédiatement placé à gauche du mot recherché



Le Niveau 1 étant coché par défaut, le chercheur peut cumuler les tris en choisissant un deuxième et même un troisième niveau. Chaque niveau est coloré différemment.



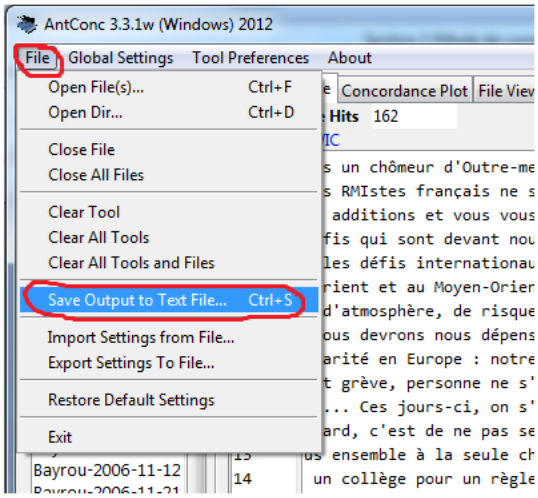
Les couleurs des niveaux de tri sont paramétrables à l'avance grâce au menu **Global Settings**.



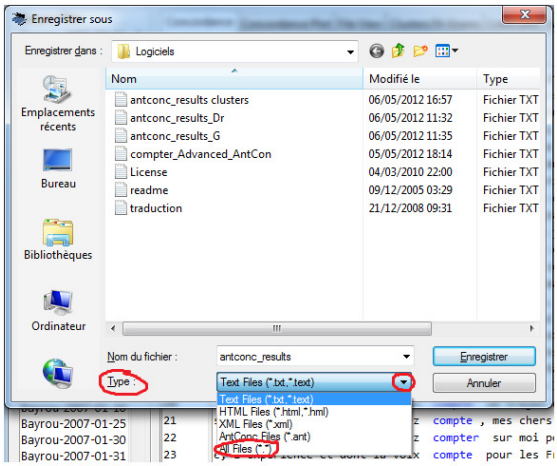
A ce stade, le résultat, visible dans la fenêtre principale, comporte trois colonnes : une colonne (**Hit**) numérote les énoncés ; une deuxième (**Kwic**) visualise les énoncés avec au centre le terme clé coloré ; une dernière colonne (**File**) mentionne, à droite de chaque énoncé, le nom du fichier source, dans le cas où la requête est lancée dans plusieurs fichiers.

D. L'enregistrement des résultats

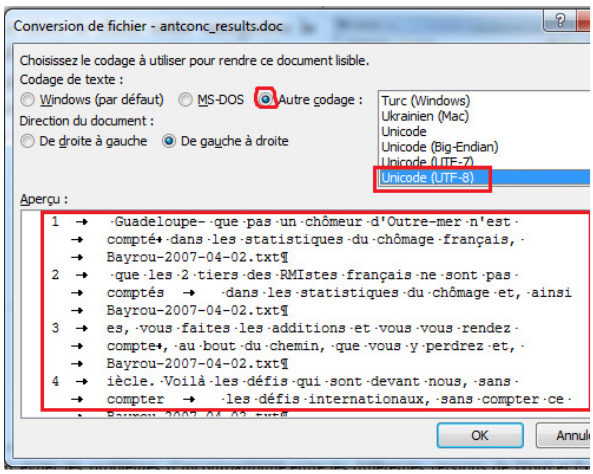
Finalement, tous les résultats qui apparaissent dans la fenêtre principale peuvent être enregistrés sur le disque local en format (.txt). Un raccourci (Ctrl+S) facile ce transfert.



On a aussi la possibilité d'enregistrer le résultat directement[21] dans WORD en spécifiant pour le type de fichier, dans le menu **Text Files**, l'option **All files (*.*).** Quant au nom du fichier, il doit comporter l'extension **(.doc**).



Lors de l'ouverture du fichier avec WORD, une fenêtre de dialogue signale que le logiciel vient de procéder au changement de l'encodage du texte. Si les énoncés visibles dans la fenêtre principale sont lisibles, la conversion peut être validée. A partir de WORD**,** le fichier obtenu peut être sélectionné dans sa totalité, copié, puis collé dans EXCEL qui reproduira les colonnes selon le réglage préalable des tabulations.

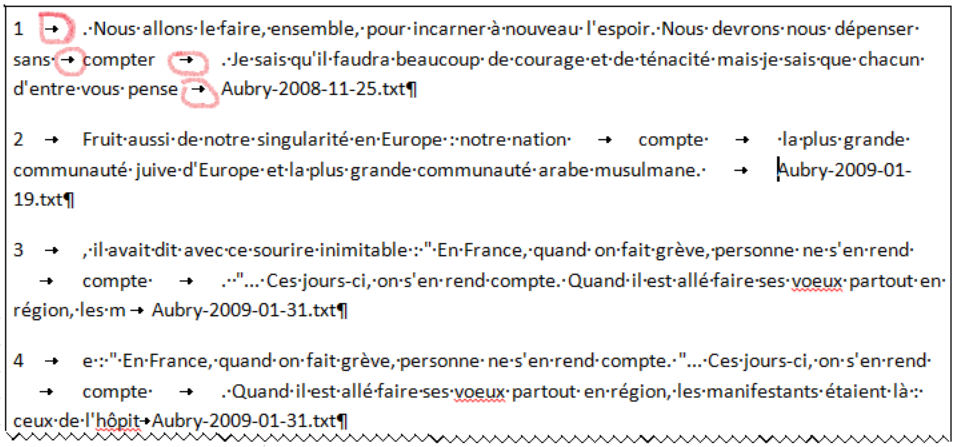


AntConc propose d'autres options de recherches pour des études statistiques comme par exemple, celle de la recherche des clusters. Nous utiliserons cette option qui permet de retrouver des suites d'unités, en fin de cette présentation, quand nous aborderons le problème des cas d'homonymie.

[21] Cette procédure fait l'économie de

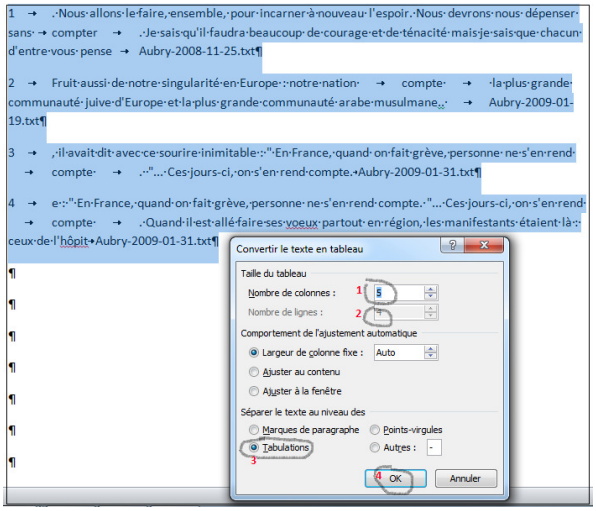
II-2.2. WORD de Microsoft Office

Transféré dans WORD, le fichier le fichier (.txt) obtenu par **AntConc** comporte suivant les tabulations placées dans les énoncés, dans chaque ligne, quatre tabulations, qui donneront cinq colonnes : une colonne pour la numérotation ; une autre pour le contexte gauche ; une troisième pour le terme-clé ; une quatrième pour le contexte droit ; et une dernière pour la référence.

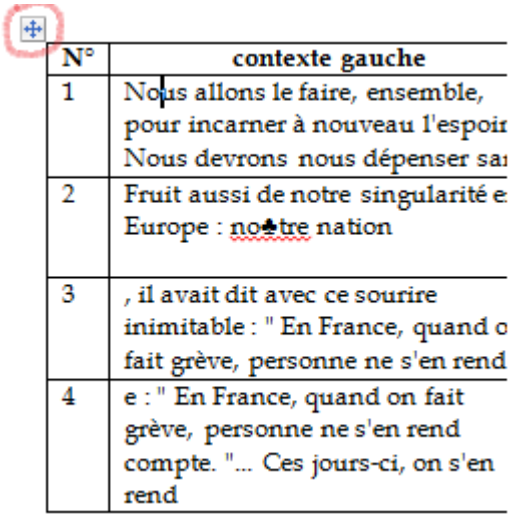

1 → .·Nous·allons·le·faire,·ensemble,·pour·incarner·à·nouveau·l'espoir.·Nous· devrons·nous· dépenser· sans· → compter → .·Je·sais·qu'il·faudra·beaucoup· de·courage·et·de·ténacité·mais·je·sais·que·chacun· d'entre·vous· pense → Aubry-2008-11-25.txt¶
2 → Fruit·aussi·de·notre·singularité·en·Europe·:·notre·nation· → compte· → ·la·plus·grande· communauté· juive·d'Europe·et·la·plus·grande·communauté·arabe·musulmane.· → Aubry-2009-01-19.txt¶
3 → ,·il·avait·dit·avec·ce·sourire·inimitable·:·"·En·France,·quand· on·fait·grève,·personne· ne·s'en·rend· → compte· → .·"...·Ces·jours-ci,·on·s'en·rend·compte.·Quand·il·est·allé·faire·ses·voeux·partout·en· région,·les·m → Aubry-2009-01-31.txt¶
4 → e·:·"·En·France,·quand·on·fait·grève,·personne· ne·s'en·rend·compte.·"...·Ces·jours-ci,·on·s'en·rend· → compte· → .·Quand·il·est·allé·faire·ses·voeux·partout· en·région,·les·manifestants·étaient·là·:· ceux·de·l'hôpit→Aubry-2009-01-31.txt¶

Pour convertir le texte en tableau, dans le menu **Insertion**, on choisit **Tableau** puis **Convertir le texte en tableau**. Une fenêtre de dialogue s'affiche pour les réglages.

Dans la fenêtre de dialogue, on vérifie le nombre de colonnes (1) que le Traitement de textes WORD va créer ainsi que le nombre de lignes (2), on coche l'option **Tabulation** (3) pour donner l'instruction de séparer texte selon les tabulations et enfin on valide avec **OK** (4).



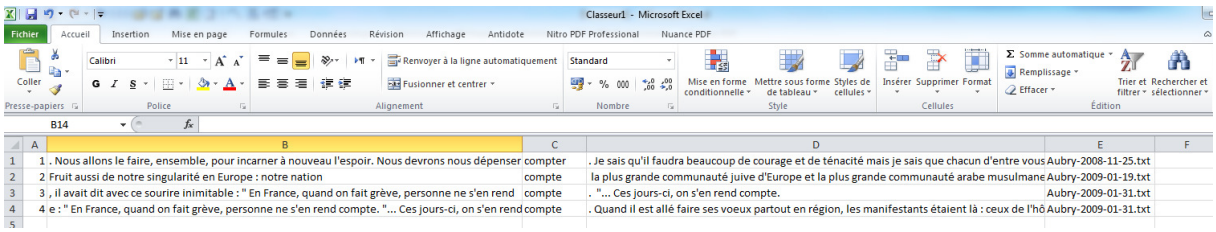
S'affiche alors le tableau, auquel on peut ajouter une ligne au-dessus pour les titres des colonnes. [Clic droit sur la petite croix bleue qui apparait quand la souris survole le tableau > **Insérer** > **Insérer des lignes au-dessus**]



N°	contexte gauche
1	Nous allons le faire, ensemble, pour incarner à nouveau l'espoir Nous devrons nous dépenser sa
2	Fruit aussi de notre singularité e Europe : notre nation
3	, il avait dit avec ce sourire inimitable : " En France, quand o fait grève, personne ne s'en rend
4	e : " En France, quand on fait grève, personne ne s'en rend compte. "... Ces jours-ci, on s'en rend

N°	contexte gauche	V/N	contexte droit	référence
1	Nous allons le faire, ensemble, pour incarner à nouveau l'espoir. Nous devrons nous dépenser sans	compter	. Je sais qu'il faudra beaucoup de courage et de ténacité, mais je sais que chacun d'entre vous pense	Aubry-2008-11-25.txt
2	Fruit aussi de notre singularité en Europe : notre nation	compte	la plus grande communauté juive d'Europe et la plus grande communauté arabe musulmane..	Aubry-2009-01-19.txt
3	, il avait dit avec ce sourire inimitable : " En France, quand on fait grève, personne ne s'en rend	compte	. "... Ces jours-ci, on s'en rend compte.	Aubry-2009-01-31.txt
4	e : " En France, quand on fait grève, personne ne s'en rend compte. "... Ces jours-ci, on s'en rend	compte	. Quand il est allé faire ses voeux partout en région, les manifestants étaient là : ceux de l'hôpital	Aubry-2009-01-31.txt

Quand on copie le résultat obtenu par **AntConc** directement dans une nouvelle feuille EXCEL, on obtient un tableau avec autant de colonnes que dans WORD.

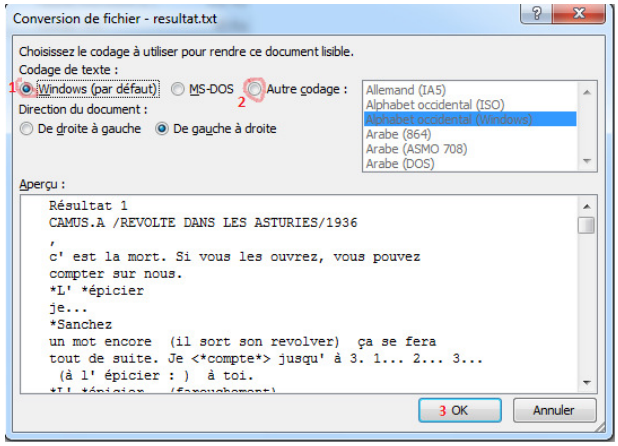


	A	B	C	D	E	F
1	1	. Nous allons le faire, ensemble, pour incarner à nouveau l'espoir. Nous devrons nous dépenser	compter	. Je sais qu'il faudra beaucoup de courage et de ténacité mais je sais que chacun d'entre vous	Aubry-2008-11-25.txt	
2	2	Fruit aussi de notre singularité en Europe : notre nation	compte	la plus grande communauté juive d'Europe et la plus grande communauté arabe musulmane	Aubry-2009-01-19.txt	
3	3	, il avait dit avec ce sourire inimitable : " En France, quand on fait grève, personne ne s'en rend	compte	. "... Ces jours-ci, on s'en rend compte.	Aubry-2009-01-31.txt	
4	4	e : " En France, quand on fait grève, personne ne s'en rend compte. "... Ces jours-ci, on s'en rend	compte	. Quand il est allé faire ses voeux partout en région, les manifestants étaient là : ceux de l'hô	Aubry-2009-01-31.txt	
5						

Attention :

Lors de l'ouverture ou de l'enregistrement de fichiers, le codage de texte peut poser un problème.

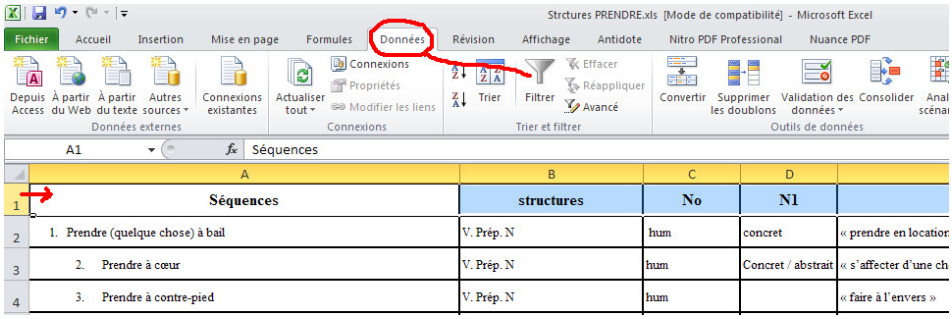
En règle générale, l'ouverture des fichiers texte est simple et ne nécessite aucune connaissance des détails sous-jacents relatifs au stockage du texte. Cependant, si les fichiers texte sont partagés avec des personnes qui utilisent d'autres langues ou d'autres codages, ou encore si ces fichiers sont téléchargés d'Internet ou créés avec d'autres systèmes informatiques, l'éditeur de textes (WORD, par exemple), lors de l'ouverture ou de l'enregistrement d'un fichier, ne pouvant pas en lire le codage, ouvre une fenêtre de dialogue permettant de choisir (2) un **autre codage**.



II-2.3. EXCEL de Microsoft Office

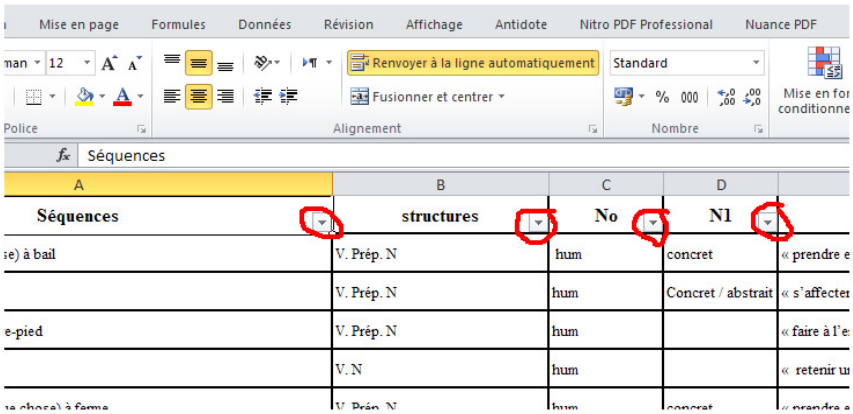
A. Le tri avec EXCEL

EXCEL de Microsoft Office[22] est un puissant programme de tri. Pour l'utiliser sur des données textuelles, on doit au préalable copier le résultat par AntConc sous forme de fichier *.txt dans un nouveau classeur EXCEL en vérifiant que les colonnes sont correctement copiées : une colonne pour le contexte gauche, une autre pour l'unité à analyser, une troisième pour le contexte droit et éventuellement une dernière pour le fichier source du corpus. Les lignes doivent contenir chacune un seul énoncé , celle qui est encadrée par les tabulations. sélectionne la première ligne (1), celle des titres des colonnes, et dans l'onglet **Données**, on choisit l'icone **Filtrer.**



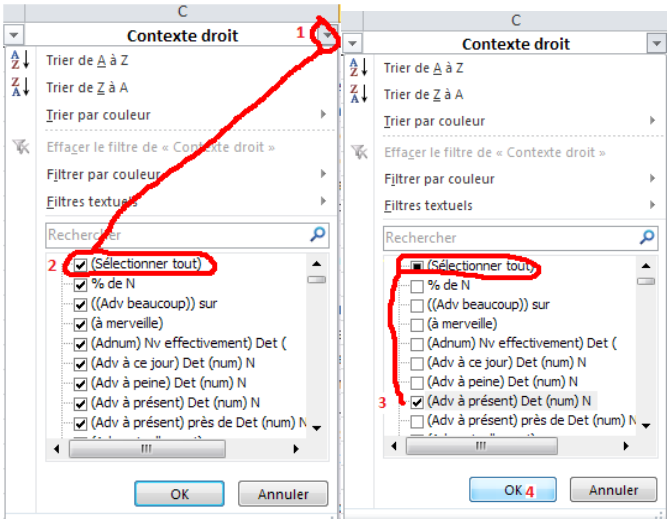
EXCEL ajoute dans le coin droit de chaque cellule de la première ligne un petit triangle cliquable qui donne accès à plusieurs possibilités de tri des données de chaque colonne du tableau selon le contenu d'une cellule au choix d'une colonne donnée.

[22] Dans notre démonstration, nous utilisons la version 2010.

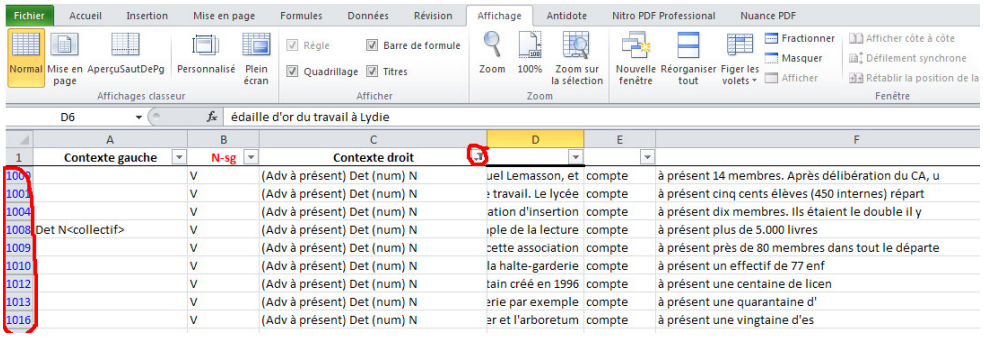


Les possibilités de tri sont visibles quand on appuie sur le petit triangle qui correpond à l'une des colonnes. Une fenêtre donne à voir sous forme de cases cochées par défaut les différents contenus des cellules de la colonne en question.

On décoche (2) d'abord la case (Sélectionner tout) pour libérer toutes les autres cases puis (3)avec l'ascenseur on fait défiler la liste des différentes possibilités et on coche la ou les cases voulues, enfin (4) on clique sur OK



Instantanément, le tableau change et ne restent visibles que les lignes qui contiennent le contenu qui a été choisi dans l'étape précédente. Le petit triangle noir est maintenant plus petit et est accompagné du symbole du tri (un entonnoir). Un autre tri peut être effectué sur le résultat obtenu par le premier tri.



Attention :

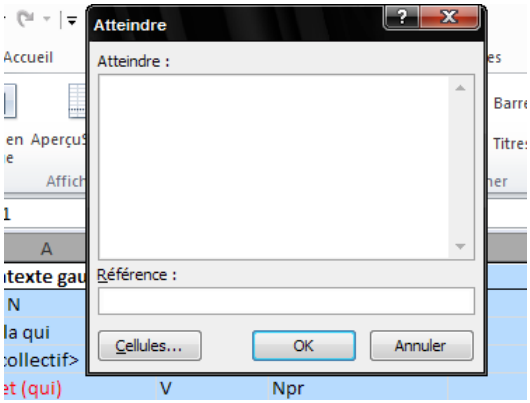
1. Ne pas oublier après chaque tri de retourner à **Sélectionner tout** pour retrouver la totalité des lignes du tableau.

2. Si le tri donne des résultats erronés, vérifier s'il n'y a pas dans le tableau une ligne vide ; EXCEL considère une ligne vide comme la fin du tableau et ne trie pas les lignes suivantes.

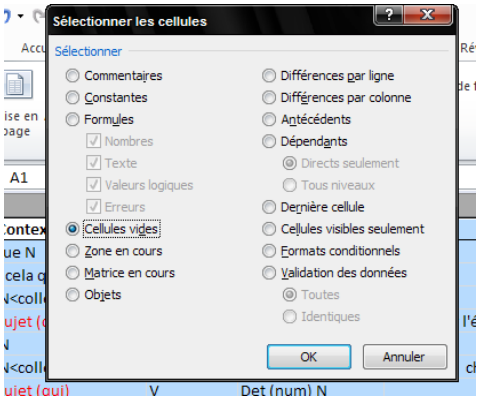
B. La suppression des lignes vides

Si, comme nous venons de le signaler, le tableau comporte des lignes vides, le tri s'arrête à la première ligne vide. Il faut donc veiller à éliminer ces lignes vides. Lorsque le tableau est un peu long, la suppression manuelle est fastidieuse. EXCEL permet de les supprimer automatiquement, ce qui est très pratique.

1) Dans EXCEL, sélectionner les colonnes du tableau. Puis presser la touche **F5** et cliquer sur le bouton **Cellules**...



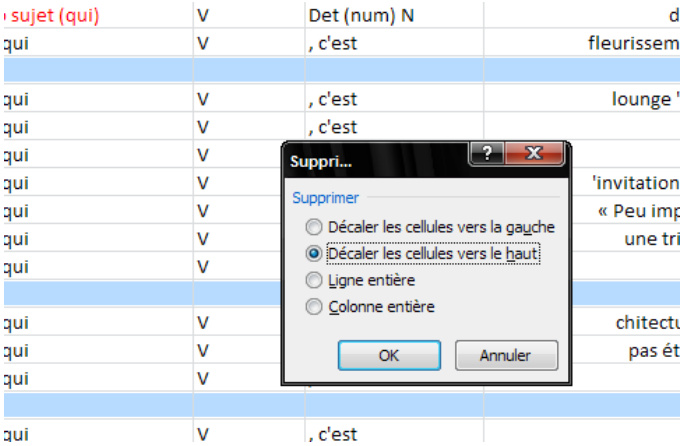
2) Sélectionner alors l'option **Cellules vides** et valider par **OK.**



3) Les lignes vides sont alors sélectionnées.

	A	B	C	D	E	F
1	Contexte gauche	N-sg	Contexte droit			
2	chaque N	V	∅	Chaque point	compte	!
3	c'est cela qui	V	∅	Et c'est cela qui	compte	!
4	Det N<collectif>	V	Det (num) N	ain Jacquot ; trésorier : Philippe André. Le club	compte	, à ce jour, onze membres et garde ses rangs, bien
5	pro sujet (qui)	V	Npr	r l'équipe départementale des Vosges (ETD 88) qui	compte	, Alain Lemare (club stéphanois), William Ratto (U
6	Det N	V	Det (num) N	édaille d'or du travail à Lydie	compte	, Anne-Claude Deroze, Marie-Noëlle Eberhard, Joël
7	Det N<collectif>	V	Det (num) N	chiffres, il est bon de savoir que Champigneulles	compte	, au dernier recensement, 7.2
8	pro sujet (qui)	V	Det (num) N	t le Territoire de Belfort, qui	compte	, au total, 9.954 adhérents. Tables rondes D'autre
9	pro sujet (qui)	V	Det (num) N	développer une association solide, qui	compte	, autour de ses animateurs, une
10	ce qui	V	, c'est	fleurissement, c'est un investissement ! - Ce qui	compte	, ce ne sont pas les « kilos de fleurs ». Ce sont
11						
12	ce qui	V	, c'est	lounge " after hours " fait un tabac. Ce qui qui	compte	, ce n'est pas la carte mais
13	ce qui	V	, c'est	at à manger froid : « Ce qui	compte	, ce n'est pas la présence de Santini sur le banc
14	ce qui	V	, c'est	s en petit comité : « Ce qui	compte	, ce n'est pas le nombre, mais le résultat final »
15	ce qui	V	, c'est	'invitation au restaurant. Mais, en amour, ce qui	compte	, ce sont les mots pour le dire. Mieux : pour l'éc
16	ce qui	V	, c'est	« Peu importe qui a l'idée, qui l'exécute, ce qui	compte	, c'est ce qui est fait. » Répondant à Jacques Lam
17	ce qui	V	, c'est	une triste idée du suffrage universel. Ce qui	compte	, c'est ce qui sortira des urnes
18	ce qui	V	, c'est	oire des deux orphelins. Ce qui	compte	, c'est ce qu'on voit avec le cœur. On a fait ensu
19						
20	ce qui	V	, c'est	chitectural. Sur le plan de l'urbanisme, ce qui	compte	, c'est davantage la qualité. Il est
21	ce qui	V	, c'est	pas été parfaite, et elle le sait... Mais ce qui	compte	, c'est d'avoir tirer les leç
22	ce qui	V	, c'est	n'arrête pas de s'interroger : «Ce qui	compte	, c'est de douter de tout», a
23						
24	ce qui	V	, c'est	forcément l'essentiel. « Ce qui	compte	, c'est de se sentir bien chez soi. » Et
25	ce qui	V	c'est	est un peu comme les restaurants de stars. Ce qui	compte	c'est de voir le patron. De

4) Cliquer avec le bouton droit de la souris sur une cellule d'une des lignes vides et choisir la commande **Supprimer**. Puis sélectionner l'option **Décaler les cellules vers le haut** et cliquer sur **OK**.



5) Il n'y a désormais plus de ligne vide dans le tableau.

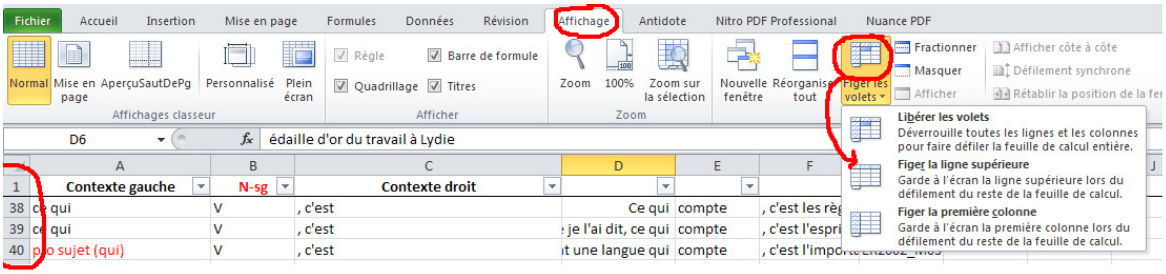
5	Det N	V	Det (num) N	chiffres, il est bon de savoir que Champigneulles	compte	, au dernier recensement, 7.2
7	Det N<collectif>	V	Det (num) N	t le Territoire de Belfort, qui	compte	, au total, 9.954 adhérents. Table
8	pro sujet (qui)	V	Det (num) N	développer une association solide, qui	compte	, autour de ses animateurs, une
9	pro sujet (qui)	V	Det (num) N	fleurissement, c'est un investissement ! - Ce qui	compte	, ce ne sont pas les « kilos de fleu
10	ce qui	V	, c'est	lounge " after hours " fait un tabac. Ce qui qui	compte	, ce n'est pas la carte mais
11	ce qui	V	, c'est	at à manger froid : « Ce qui	compte	, ce n'est pas la présence de Sant
12	ce qui	V	, c'est	s en petit comité : « Ce qui	compte	, ce n'est pas le nombre, mais le
13	ce qui	V	, c'est	'invitation au restaurant. Mais, en amour, ce qui	compte	, ce sont les mots pour le dire. M
14	ce qui	V	, c'est	« Peu importe qui a l'idée, qui l'exécute, ce qui	compte	, c'est ce qui est fait. » Répondan
15	ce qui	V	, c'est	une triste idée du suffrage universel. Ce qui	compte	, c'est ce qui sortira des urnes
16	ce qui	V	, c'est	oire des deux orphelins. Ce qui	compte	, c'est ce qu'on voit avec le cœur.
17	ce qui	V	, c'est	chitectural. Sur le plan de l'urbanisme, ce qui	compte	, c'est davantage la qualité. Il est
18	ce qui	V	, c'est	pas été parfaite, et elle le sait... Mais ce qui	compte	, c'est d'avoir tirer les leç
19	ce qui	V	, c'est	n'arrête pas de s'interroger : «Ce qui	compte	, c'est de douter de tout», a
20	ce qui	V	, c'est	forcément l'essentiel. « Ce qui	compte	, c'est de se sentir bien chez soi.
21	ce qui	V	, c'est	est un peu comme les restaurants de stars. Ce qui	compte	, c'est de voir le patron. De
22	ce qui	V	, c'est	ment ? - Je ne préfère pas parler d'écart. Ce qui	compte	, c'est déjà les battre. Après
23	ce qui	V	, c'est	on joue à Pau-Orthez ou contre Strasbourg, ce qui	compte	, c'est d'essayer de gagner contre
24	ce qui	V	, c'est	, fidèlement installées tribune Hazotte. « Ce qui	compte	, c'est jouer. Même si, avec l'ASN
25	ce qui	V	, c'est	, fidèlement installées tribune Hazotte. « Ce qui	compte	, c'est jouer. Même si, avec l'ASN

C. Le figement des volets des titres

Lorsqu'on fait défiler un tableau verticalement ou horizontalement, souvent si celui-ci est long, on ne peut plus voir la ligne qui contient les titres des

colonnes ou la colonne qui contient les titres des lignes. Pourtant celles-ci sont indispensables pour saisir les informations dans la bonne colonne ou tout simplement pour savoir à quelle colonne ou ligne correspond une cellule donnée.

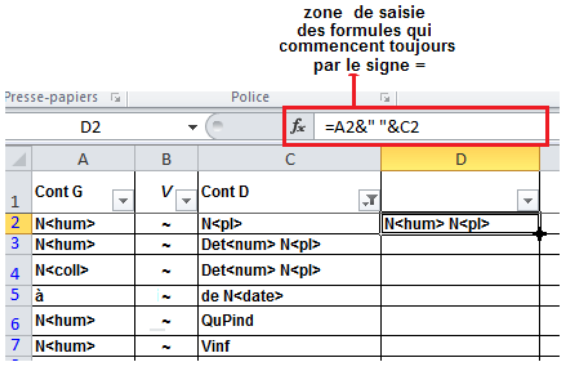
EXCEL permet de figer la ligne ou la colonne des titres. Cliquer sur la cellule A1 (celle qui comporte les titres des colonnes) et choisir dans le menu **Affichage** l'option **Figer les volets**. On a alors le choix entre figer la ligne des titres, ou figer une ou plusieurs colonnes à la fois ou figer tous les volets. Désormais, en se déplaçant dans le tableau de haut en bas ou de gauche à droite, ce qu'on a décidé de figer reste visible. Pour libérer le tableau, refaire les mêmes opérations.



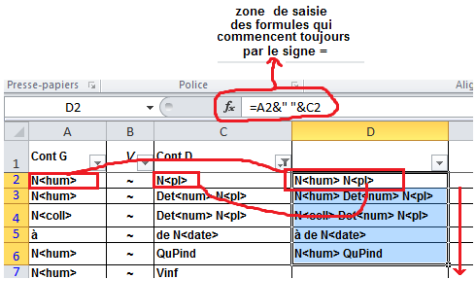
D. La fusion de deux colonnes

On peut concaténer deux colonnes en utilisant une formule EXCEL avec l'opérateur **&**. Par exemple, si les deux colonnes à unir sont **A** et **C**, il suffit de cliquer dans une cellule d'une colonne vide (dans notre exemple **D**) puis utiliser, dans la zone de saisie des formules, la commande **=Ax&"●"&Cx** (où x est le numéro de la ligne dans laquelle se trouve la cellule (ici D2) et le point une espace ; l'ordre C A ou A C, dépend du choix de l'utilisateur).

La formule **=A2&"●"&C2**, pour la ligne 2 donne le contenu de la cellule A suivi du contenu de la cellule C. On peut, si on le désire, inverser l'ordre des colonnes en inversant l'ordre des lettres des deux colonnes =C2&"●"&A2 donnera le contenu de la cellule C suivi du contenu de la cellule A.



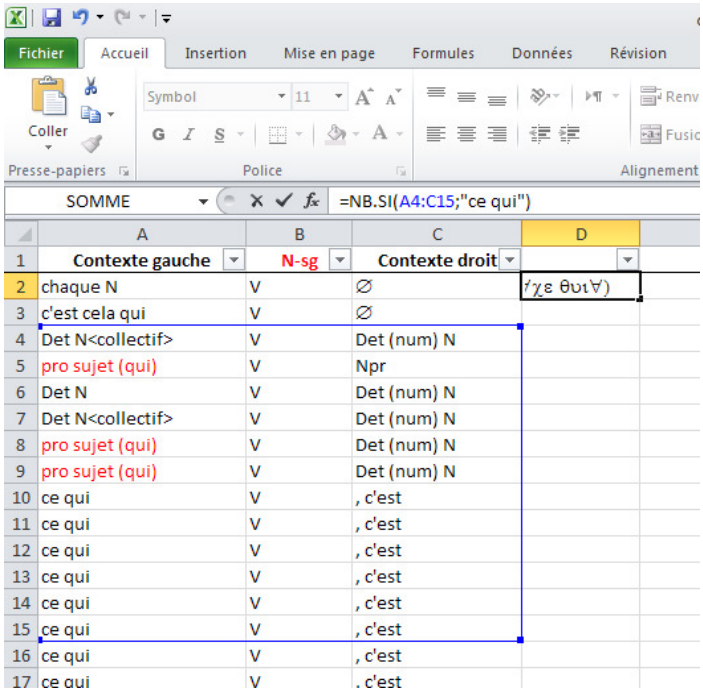
Quand le résultat est obtenu, pour appliquer la même formule à la colonne il suffit alors d'accrocher la petite croix noire qui apparait dans le coin gauche en bas de la cellule du résultat de la formule et de tires vers le bas jusqu'à la ligne voulue.



E. Le comptage des cellules contenant une chaîne de caractères

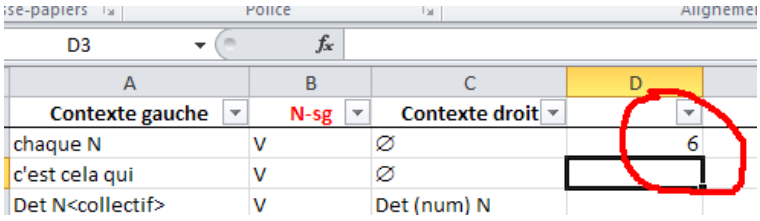
1) Pour compter le nombre des cellules contenant une chaîne de caractères précise (par exemple **ce qui**) dans une plage de cellules (ici A4:C15), cliquer dans une cellule vide (ici D2) puis saisir dans la zone des formules la formule suivante :

=NB.SI(A4:C15;"ce qui")



	A	B	C	D
1	Contexte gauche	N-sg	Contexte droit	
2	chaque N	V	∅	
3	c'est cela qui	V	∅	
4	Det N<collectif>	V	Det (num) N	
5	pro sujet (qui)	V	Npr	
6	Det N	V	Det (num) N	
7	Det N<collectif>	V	Det (num) N	
8	pro sujet (qui)	V	Det (num) N	
9	pro sujet (qui)	V	Det (num) N	
10	ce qui	V	, c'est	
11	ce qui	V	, c'est	
12	ce qui	V	, c'est	
13	ce qui	V	, c'est	
14	ce qui	V	, c'est	
15	ce qui	V	, c'est	
16	ce qui	V	, c'est	
17	ce qui	V	, c'est	

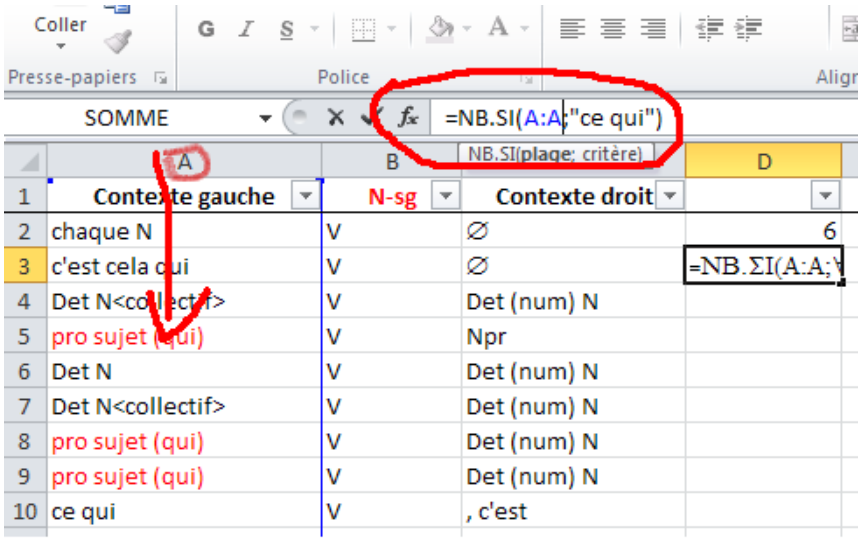
Immédiatement apparait le nombre des occurrences demandées



A	B	C	D
Contexte gauche	N-sg	Contexte droit	
chaque N	V	∅	6
c'est cela qui	V	∅	
Det N<collectif>	V	Det (num) N	

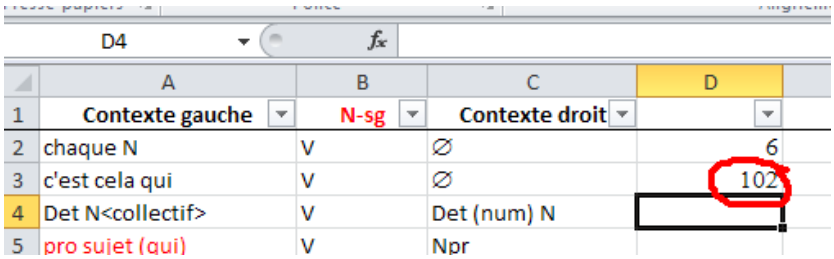
2) Pour compter les cellules contenant une chaîne de caractères précise (par exemple **ce qui**) dans une seule colonne (ici la colonne A), saisir la formule suivante :

=NB.SI(A:A;"ce qui")



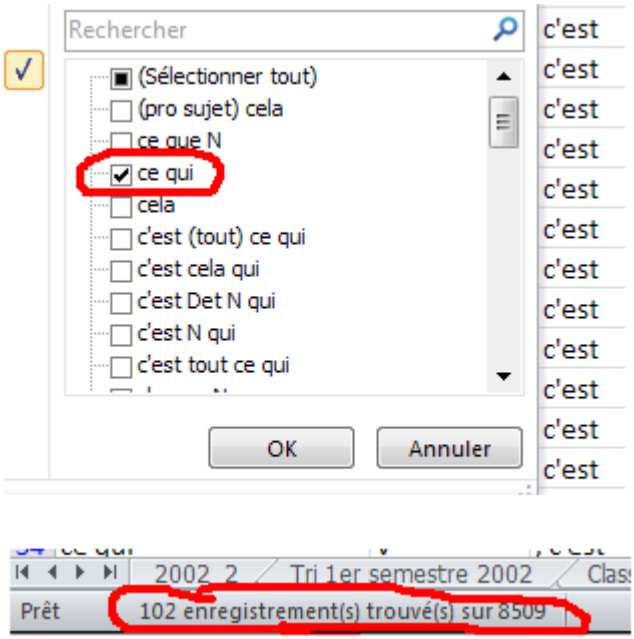
	A	B	C	D
1	Contexte gauche	N-sg	Contexte droit	
2	chaque N	V	∅	6
3	c'est cela qui	V	∅	=NB.SI(A:A;"
4	Det N<collectif>	V	Det (num) N	
5	pro sujet (qui)	V	Npr	
6	Det N	V	Det (num) N	
7	Det N<collectif>	V	Det (num) N	
8	pro sujet (qui)	V	Det (num) N	
9	pro sujet (qui)	V	Det (num) N	
10	ce qui	V	, c'est	

On obtient rapidement



	A	B	C	D
1	Contexte gauche	N-sg	Contexte droit	
2	chaque N	V	∅	6
3	c'est cela qui	V	∅	102
4	Det N<collectif>	V	Det (num) N	
5	pro sujet (qui)	V	Npr	

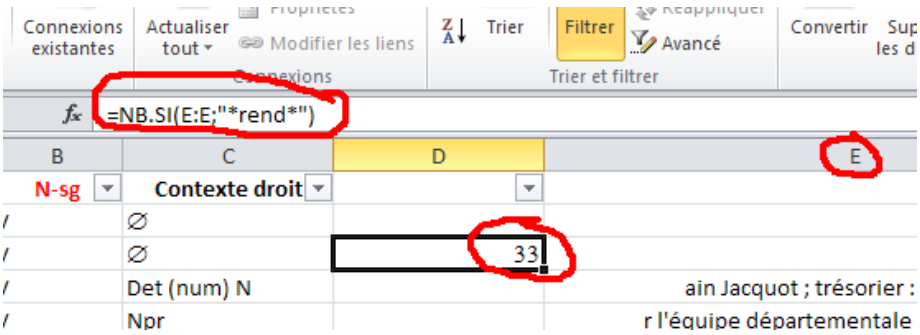
Le résultat peut se vérifier avec l'opération de tri[23]



3) des caractères génériques peuvent être utilisés pour remplacer un ou plusieurs caractères. La formule suivante comptabilise toutes les cellules qui contiennent l'expression **rend** dans la colonne **E**.

[23] Voir plus haut, §**A. Le tri avec** EXCEL**.**

=NB.SI(C:C;"*rend*")



contiennent une chaîne de quatre caractères (les mots de quatre lettres) :

=NB.SI(C:C;"????")

II-2.4. **Combinaison AntConc / EXCEL** [Cette section peut être sautée en première lecture]

Après l'élimination des cas « parasites », les résultats obtenus par **AntConc** posent encore le problème des homonymes. Ainsi, sur le plan morphologique, la forme ***compte***, en elle-même, est ambiguë. Seul l'environnement phrastique permet de lever cette ambiguïté. Même les concordanciers catégoriseurs, tels que Unitex[24] ou TXM[25], ne sont pas encore capables dans l'état actuel des recherches en traitement automatique du langage de lever de telles ambiguïtés. Seuls le tri manuel, ou plutôt le travail humain permettent d'en venir à bout.

En effet, la forme *compte* correspond au verbe *compter* à la première et à la troisième personne du singulier du présent de l'indicatif et du subjonctif, mais également au nom comme dans « compte-rendu » ou « compte tenu » ou encore « tenir compte ». La forme *comptes* correspond au verbe *compter* à la deuxième personne du singulier du présent de l'indicatif et du subjonctif, mais aussi au pluriel du nom *compte* comme dans « la Cour des comptes ». La forme *comptant* correspond au participe présent du verbe *compter* ainsi qu'au nom ou à l'adjectif « payer au comptant » ou « argent comptant ».

Pour notre part, nous avons procédé de la manière suivante en effectuant un va-et-vient entre le concordancier **AntConc** et l'éditeur EXCEL. Le premier permet de relever les fréquences des structures itératives, et le second d'éliminer

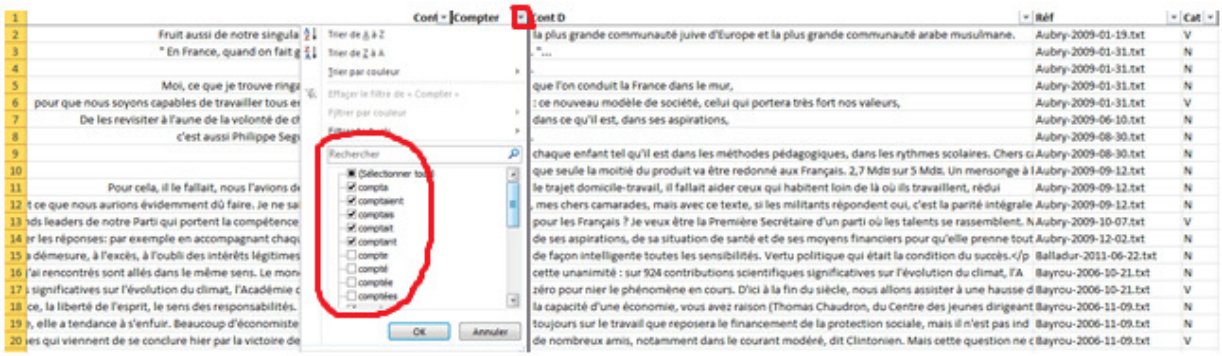
[24] Unitex est un ensemble de logiciels permettant de traiter des textes en langues naturelles en utilisant des ressources linguistiques. Ces ressources se présentent sous la forme de dictionnaires électroniques, de grammaires et de tables de lexique-grammaire. Elles sont issues de travaux initiés sur le français par Maurice Gross au *Laboratoire d'Automatique Documentaire et Linguistique* (LADL).
Adresse = http://www-igm.univ-mlv.fr/~unitex/

[25] TXM est une plate-forme open source d'analyse textométrique conçue grâce à la collaboration de l'ENS de Lyon, le CNRS, l'INRP, l'Université de Lyon 2, l'Université de Franche-Comté, l'Université de Nice Sophia Antipolis et l'Université de Paris 3. Elle est livrée avec quatre corpus étiquetés morphosyntaxiquement et permet d'automatiser l'étiquetage et la lemmatisation d'autres corpus au cours du processus d'importation. Adresse = http://textometrie.ens-lyon.fr/

les cas des formes nominales, de classer les différentes structures du verbe *compter* selon les synonymes et de faire les statistiques.

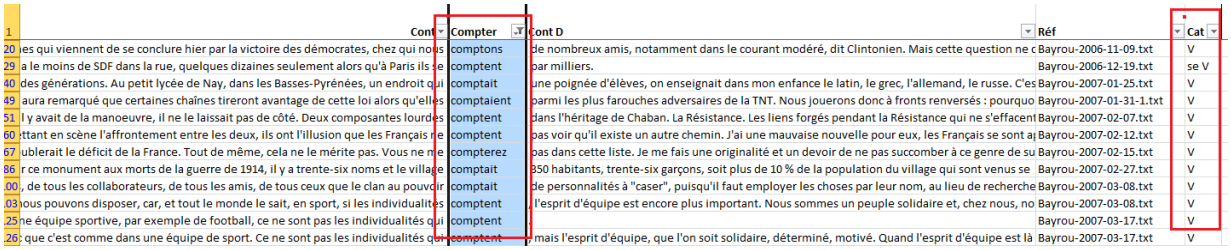
Dans un premier temps, avec le simple filtre des données que nous avons placé au niveau de la ligne des titres, nous isolons les cas sans ambiguïté des formes verbales, à savoir :

> compter, comptaient, comptais, comptait, comptai, comptâmes, comptassent, comptasses, comptasse, comptassiez, comptassions, comptas, comptâtes, comptât, compta, comptent, compteraient, compterait, compterai, compteras, comptera, comptèrent, compterez, compteriez, compterions, compterons, compteront, comptez, comptiez, comptions, comptions, comptons, compterais



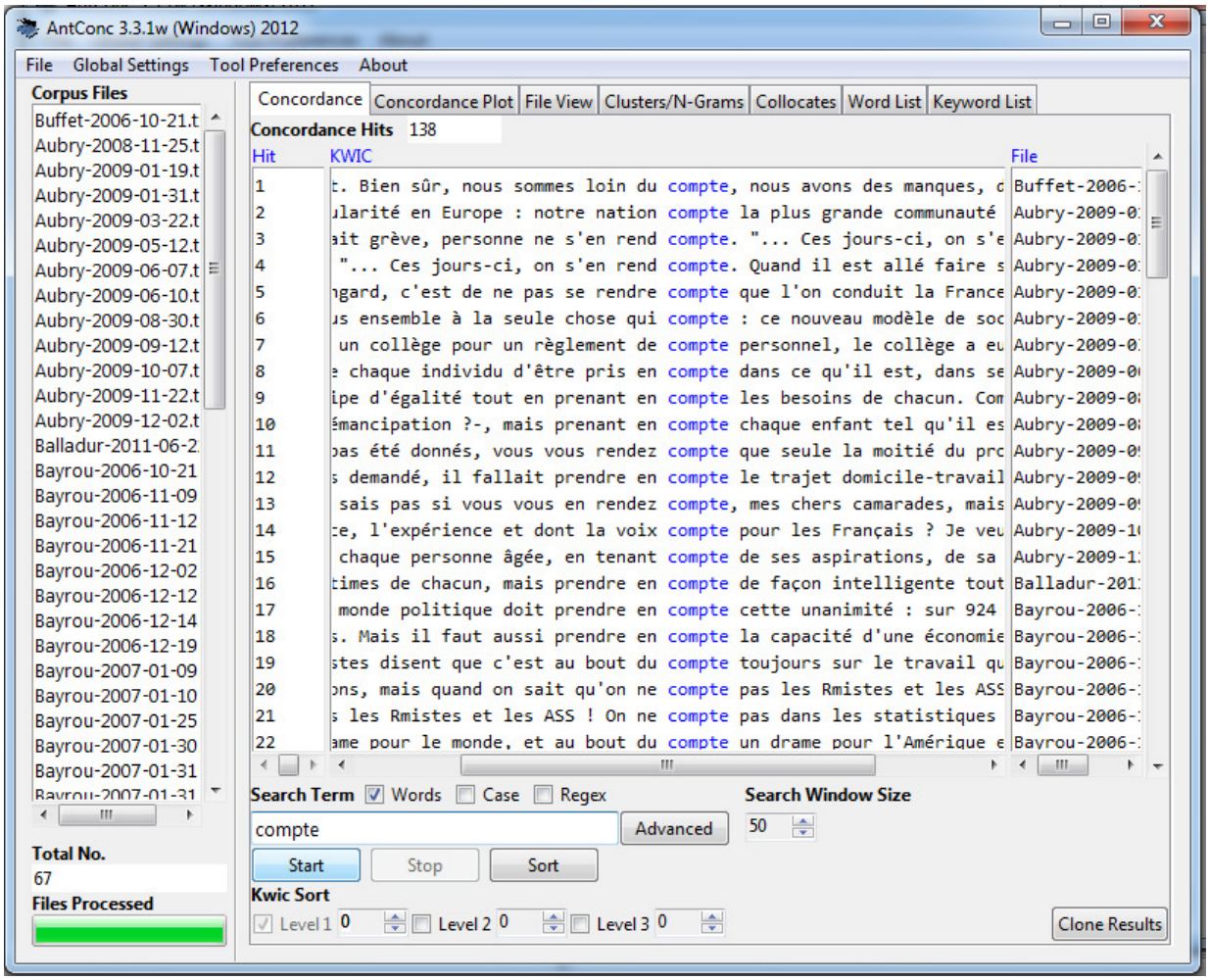
Nous gardons provisoirement de côté les cas de *compte*, *comptes*, *compté*, *comptés*, *comptée*, *comptées*, *comptant* qui peuvent prêter à confusion.

Dans la colonne prévue pour la catégorie, nous inscrivons un V pour marquer les cas des formes verbales



Notre corpus littéraire présente 7 occurrences de (*tu*) *comptes* sur 200 occurrences du verbe *compter*. Quant aux deux autres corpus journalistique et politique ils n'en comptent (!) aucune occurrence, ce qui n'est pas pour surprendre puisqu'en discours journalistique une forme de 1^{e} personne du présent (de l'indicatif ou du subjonctif) ne peut guère se présenter que dans une citation et dans les discours politiques l'orateur s'adresse à un auditoire qu'il ne peut pas tutoyer.

En ce qui concerne le singulier *compte*, nous relevons 2004 cas (verbe et nom) sur les 3508 concordances que comporte le corpus du Discours politique, toutes formes confondues.

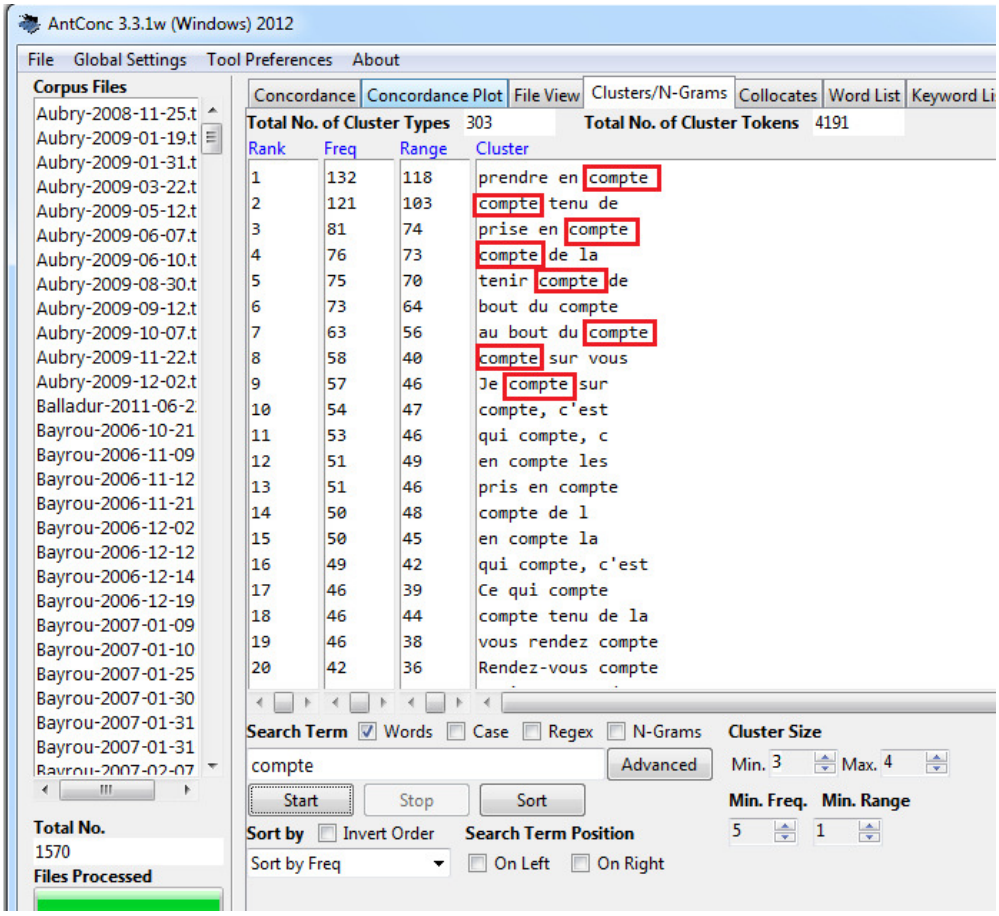


Pour isoler les occurrences verbales de celles des noms, nous faisons intervenir l'option **Clusters**[26] dans **AntConc**. Cette option permet, selon le réglage du nombre des termes voulu (**Cluster Size**), d'obtenir une liste de séquences contenant le terme recherché.

Dans notre exemple, une fois le corpus chargé, nous avons lancé la requête avec la forme simple ***compte***. L'onglet **Clusters/N-Grams** de la fenêtre principale (1) offre un tableau avec quatre colonnes :

- Rank « numéro » ;
- Freq « fréquence de la séquence » ;
- Range « classement » ;
- Cluster « la séquence relevée ».

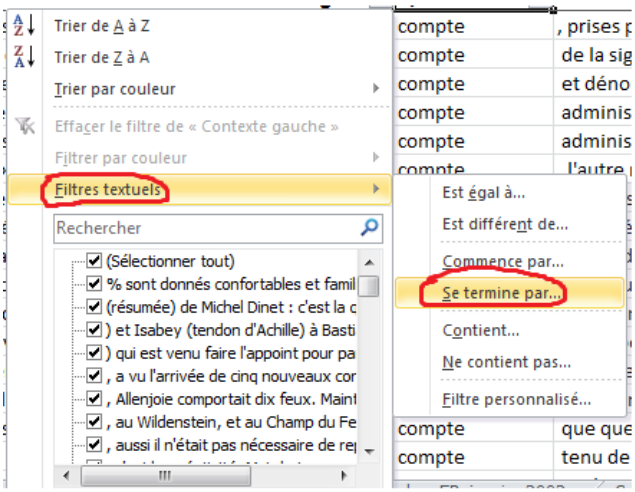
[26] Dans la dernière version, cette option est augmentée de sa variante N-Grams.



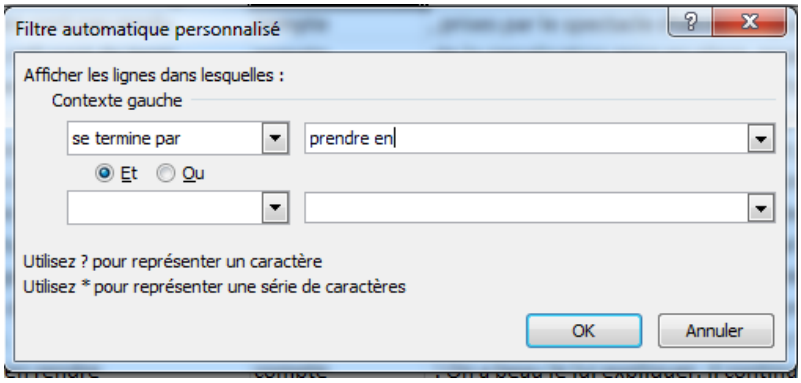
Sans les réglages de la position du terme recherché, les résultats montrent, tout de même, les structures récurrentes et leurs fréquences respectives. Nous pouvons savoir, par exemple, que le corpus comporte 123 occurrences de la structure « prendre en compte » où la forme *compte* est un nom. Il suffit alors d'ouvrir le fichier EXCEL (ici du corpus discours politique) et, avec le tri, de rechercher dans un premier temps les occurrences de la forme *compte*.

	Contexte gauc	V/N	Contexte droit	V/N
2	e mais les personnes âgées ne s'en sont pas rendu	compte	, prises par le spectacle d'un remarquable couple	
3	mme celui de ces automobilistes refusant de tenir	compte	de la signalisation mise en place, essayant de pa	
4	ue deux années de travail ne soient pas prises en	compte	et dénonçant les incohérences de l'Etat. Excep	
5	ion de l'année. A l'ordre du jour : l'examen du	compte	administratif 2000 et le vote du budget primitif	
6	hérents. Dans de telles conditions, l'examen du	compte	administratif comme celui du budget primitif se s	
10	déboussolés » L'un n'arrivait pas à trouver son	compte	, l'autre manquait de « cents », un autre encore c	
17	nes. Parmi ses 4.500 habitants, le Val de Venne	compte	300 personnes inscrites à l'Amicale des donneurs	
18	déjà prononcée pour cette structure. En tenant	compte	de ces éléments, le conseil municipal a décidé d'	
20	c une réelle satisfaction. <div_type="article">	Compte	-rendu du conseil municipal Il est précisé que l	
21	riage et réunion de famille après des obsèques.	Compte	-tenu du passage aux 35 heures, le contrat de trav	
24	courant 2001, c'est encourageant. Pour 2002, je	compte	prochainement prendre contact avec les gérants de	
25	-il ? Il est arrivé à ses fins sans s'en rendre	compte	! On a beau le lui expliquer, il continue de dout	
28	s de motif avéré d'en décider autrement. » « Je	compte	sur le sens de l'intérêt général et du service pu	
35	Aujourd'hui, il y a 22.000 habitants et la mairie	compte	600 agents. Je voulais donc retrouver quelque cho	
53	kets ! « Depuis l'ouverture, on n'a pu prendre en	compte	que quelques paris et puis tout s'est bloqué », e	
57	facturés 300 F l'unité. Car tout bien réfléchi,	compte	tenu de la lenteur avec laquelle les flots ont co	

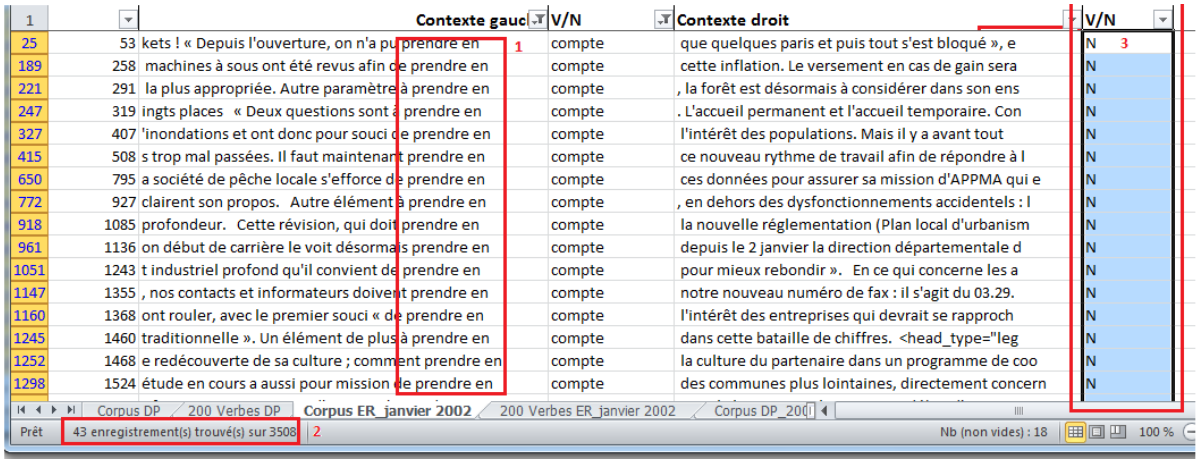
Dans un deuxième temps, nous recherchons, dans la colonne du contexte gauche, grâce aux **Filtres textuels** et **Se termine par...** la séquence « prendre en ».



La séquence « prendre en » est saisie comme dans la capture d'écran suivante.



Le logiciel relève 43 cas (2) que nous notons N (pour nom) dans la colonne prévue (3) pour noter les formes nominales à éliminer.



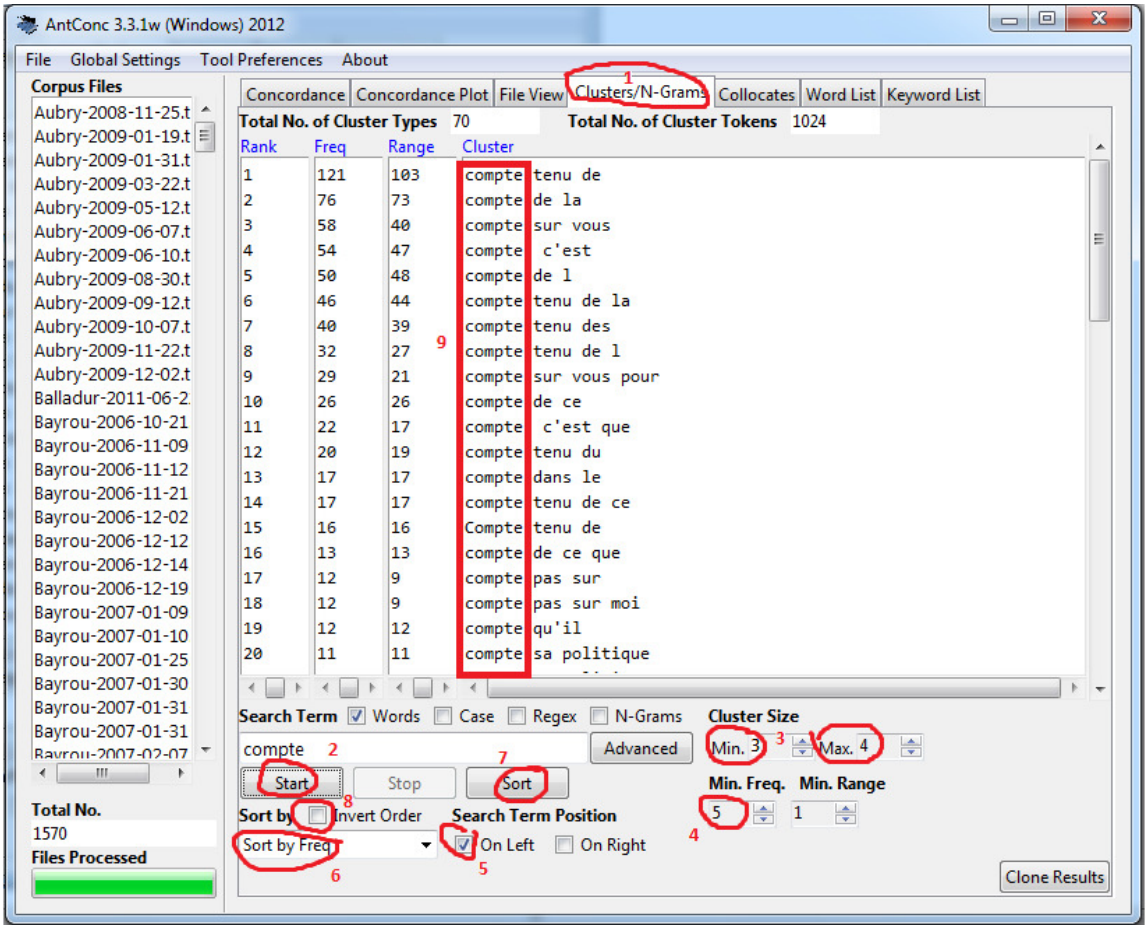
De retour à **AntConc**, nous pouvons affiner les tris selon plusieurs paramètres.

Le réglage **Cluster Size** (3) donne à choisir le nombre de termes y compris le terme recherché. La valeur Min.3 et Max.4 signifie qu'on demande au programme de relever toutes les séquences composées au minimum de trois mots et au maximum de quatre mots comprenant la forme *compte*. Dans la version actuelle, si la recherche comporte plus d'un mot, il faut prévoir d'ajouter

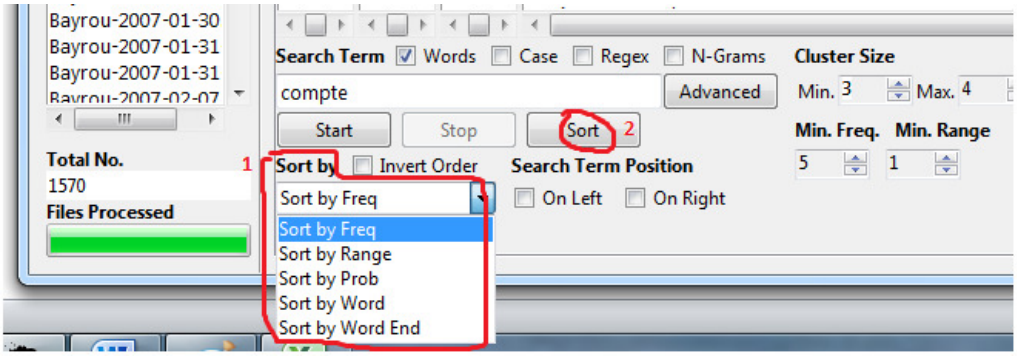
le nombre des mots recherchés aux valeurs minimale et maximale, sinon les résultats seraient inexploitables.

D'autre part, nous avons choisi la valeur 5 pour la fréquence de la séquence (4) afin de ne retenir, pour les besoins de la démonstration, que les hautes fréquences. À tout moment, l'ordre des résultats peut être inversé (8).

La position du terme recherché par rapport aux autres mots de la séquence **peut** également être contrôlée grâce à **Search Term Position**. **On Left** (5) positionne, comme dans la capture, le terme recherché au début de la série, et **On Right** à gauche. Par défaut, quand aucune option n'est retenue, la position du mot-clé est dans l'ordre linéaire des énoncés.



Enfin, pour ordonner les concordances, le menu **Sort by** (1) donne à choisir entre, par défaut, un ordre par fréquence, ou bien par classement, ou encore par la probabilité du premier mot dans le groupe précédant les mots restants, ou enfin par le début ou la fin du mot. Les choix de ce menu sont validés par un clic sur le bouton **Sort** (2).



Comme pour la séquence « *prendre en compte* », nous avons répété le procédé avec des séquences du type « *compte tenu* », « *compte rendu* », etc. que nous avons éliminés comme étant des cas de formes nominales.

Ainsi, plusieurs va-et-vient étaient nécessaires pour isoler les formes verbales et commencer finalement le dépouillement et l'analyse proprement dite du corpus.

II–2.5. Conclusion

Les opérations proposées dans cette seconde section pour reproduire nos résultats exposés dans la première section et mener des analyses comparatives de corpus originales peuvent paraître complexes. Mais il faut considérer que l'effort pour suivre leur mise en place est d'un grand profit, puisqu'il débouche sur l'aptitude à comparer quantitativement deux ou trois corpus textuels de manière efficace, sans avoir besoin d'une formation en traitement automatique des langues (TAL), c'est-à-dire sans apprendre à manier un langage de programmation.

À l'heure actuelle, la linguistique française, en tant que discipline universitaire, subit une double pression, de la didactique des langues (Français Langue Maternelle ou Français Langue Etrangère) et du TAL. Ce sont essentiellement ces formations qui assurent aujourd'hui des débouchés en sciences du langage. La méthode d'analyse des corpus textuels que nous proposons dans ce Cahier vise à constituer un pont entre la linguistique française et le TAL. Par ailleurs pour les étudiants 'exophones', c'est-à-dire non francophones natifs (et nous pensons en particulier aux étudiants arabophones), cette méthode – même si elle ne prétend pas se substituer aux jugements d'acceptabilité, dans la mesure où l'absence d'une construction dans un ou plusieurs corpus ne prouve pas qu'elle est complètement inusitée en français contemporain – permet d'**EVALUER L'USAGE DES STRUCTURES LEXICO-SYNTAXIQUES** (celles qui relèvent de la syntaxe associée aux lexèmes) **DANS DIFFERENTS TYPES DE DISCOURS.**

RÉFÉRENCES

Dubois, Jean & Dubois-Charlier, Françoise (1997), *Les verbes français*, Paris : Larousse, version électronique accessible sur le site http://rali.iro.umontreal.ca/Dubois/

François, Jacques (2007), *Pour une cartographie de la polysémie verbale*. Louvain : Peeters.

François, Jacques (2008), *Les grammaires de construction – Un bâtiment ouvert aux quatre vents*. Cahier du CRISCO n°26. Université de Caen-Basse-Normandie [téléchargeable sur le site www.crisco.unicaen.fr]

François, Jacques (2010a), *L'attestation des combinaisons lexicales à l'aide de la base de données textuelles FRANTEXT*. Cahier du CRISCO n°29. Université de Caen-Basse-Normandie. [Téléchargeable sur le site www.crisco.unicaen.fr]

François, Jacques (2010b), « Le repérage des segments répétés et la polysémie lexicale (à propos des cooccurrences de *force(s)* dans trois corpus journalistiques) ». *Revue de Sémantique et Pragmatique* n°27, p.11-30. Presses Universitaires d'Orléans.

François-Sandli, Sayda (2007), *La polysémie des verbes de communication linguistique en français*. Mémoire de Master 2, université de Rouen.

Gasiglia, Nathalie (2004), « Faire coopérer deux concordanciers-analyseurs pour optimiser les extractions en corpus », *Revue française de linguistique appliquée*, IX, pp. 45- 62.

Gross, Maurice (1975), *Méthodes en syntaxe. Le régime des constructions complétives*. Paris : Hermann.

Herslund, Michael (2000), « Le participe présent comme co-verbe », *Langue Française*, No. 127, 2000, p. 86-94.

Lebart, Lebart & Salem, André (1994), *Statistique Textuelle*, Dunod, 344 p. Epuisé, accessible en ligne ► http://egsh.enst.fr/lebart/ST.html

Levin, Beth (1993), *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago : University of Chicago Press.

Mayaffre, Damon (2005), « Rôle et place des corpus en linguistique : réflexions introductives ». *Texto !* , vol. X, n° 4.

Mayaffre, Damon (2008), « De l'occurrence à l'isotopie. Les co-occurrences en lexicométrie », in Mathieu Valette (éd.), *Textes, documents numériques, corpus. Pour une science des textes instrumentée*, *Syntaxe & Sémantique*, n°9, 2008, pp. 53-72.

Pincemin Bénédicte (2006), « Concordances et concordanciers - De l'art du bon KWAC ». Soumission à *Documents numériques et interprétation - Corpus en Lettres et Sciences sociales*, Albi, 10-14 juillet 2006.

Pincemin Bénédicte (2009), « Panorama bref et pragmatique des outils de textométrie et apparentés », Fiche réalisée à l'intention des participants, Ecole thématique CNRS MISAT (Méthodes Informatiques et Statistiques en Analyse de Textes), Besançon, 15-19 juin 2009, éditée en juillet 2009. Fiche disponible en ligne ► http://icar.univ-lyon2.fr/membres/bpincemin/biblio/pincemin_misat09_logiciels.pd

Van den Einde, Karel & Mertens, Piet (2006*), Le dictionnaire de valence DICOVALENCE – Manuel d'utilisation – version 1.2* [pdf, téléchargeable sur le site bach.arts.kuleuven.be/dicovalence]

Une note pour finir :

Toute référence à un site de corpus, de concordancier, de revue etc., doit comporter, outre l'URL et l'acronyme, le nom en toutes lettres de la ressource consultée ou utilisée et la date de consultation, par exemple :

► http:// www.crisco.unicaen.fr - Centre de Recherche Inter-langues sur la Signification en COntexte (CRISCO) - Université de Caen-Basse-Normandie (consulté le 10 mai 2012)

III

Cahiers du CRISCO

ISSN 2116-6242

Cahiers de recherche linguistique diffusés par le CRISCO depuis 2000. Les ***Cahiers du CRISCO*** sont distribués gratuitement.

A. Cahiers 1-12 en format papier

n°1	Jacques FRANÇOIS, *Sur la grammaire de van Valin* (septembre 2000)
n°2	Véronique LENEPVEU, *Adjectifs et adverbes : une corrélation syntactico-sémantique* (janvier 2001)
n°3	Jacques FRANÇOIS & Hansjakob SEILER, *Gespräche über Hansjakob Seilers Universalienforschung* (mai 2001, version originale en allemand)
n°4	Denis RAMASSE, *L'intonation des phrases présentant un détachement à gauche en français : une première approche* (juin 2001)
n°5	Heike BALDAUF, *Voyage au bout de la langue : quelques remarques sur des énoncés minimaux* (mars 2002)
n°6	Jacques FRANÇOIS (avec la participation de Hansjakob SEILER, *La faculté de langage et les universaux : perspectives fonctionnalistes actuelles* (mars 2002)
n° 7	Catherine CAMUGLI GALLARDO, *La langue des jeunes en Italie — Guide à une enquête de terrain* (mars 2002)
n° 8	Jacques FRANÇOIS, *La représentation des structures verbo-nominales et du figement verbal dans deux formalismes de grammaire fonctionnelle* (mai 2002)
n° 9	Valeriano BELLOSTA von COLBE & Jacques FRANÇOIS, *Windowing of attention and the core-periphery boundary* (septembre 2002)
n° 10/1	Maxi KRAUSE, *AUS, élément prépositionnel, élément constitutif de particules diverses et particule verbale* (septembre 2002)
n° 10/2	Maxi KRAUSE, *VON, préposition et élément constitutif d'adverbes et de tournures adverbiales* (octobre 2002)
n° 10/3	Maxi KRAUSE, *BEI, préposition, élément constitutif de particules diverses et particule verbale* (novembre 2002)
n° 10/4	Maxi KRAUSE, *MIT, préposition, élément constitutif de particules diverses et particule adverbiale* (novembre 2002)
n° 11	Richard RENAULT, *Morphosyntaxe des participes finnois* (décembre 2002)
n° 12	Dominique LEGALLOIS (coordination), *Unité(s) du texte* [Journée Scientifique du CRISCO, 6 décembre 2002] (juin 2003)

B. Cahiers 13-28 téléchargeables sur le site du CRISCO : www.crisco.unicaen.fr

n°13 Jacques FRANÇOIS (dir.), *Aspects de la « Role and Reference Grammar »* [Journée Scientifique du CRISCO, 14 mars 2003] (septembre 2003)

n°14 Jacques FRANÇOIS, Jean-Luc MANGUIN & Bernard VICTORRI, *La réduction de la polysémie adjectivale en cotexte nominal : une méthode de sémantique calculatoire (octobre 2003)*

n°15 Emmanuelle ROUSSEL*, Le commentaire associé au passif anglais* (février 2004)

n°16 Corinne OZOUF*, Caractère différentiel et relation d'équivalence entre* ***voir*** *et* ***regarder*** (juin 2004)

n°17 Jean-Luc MANGUIN & Jacques FRANÇOIS (coordination), *Le Dictionnaire Electornique des Synonymes du CRISCO – Un mode d'emploi à trois niveaux* (juillet 2004)

n°18 Dominique LEGALLOIS & Song-Nim KWON*, Sémantique lexicale et examen écologique de la co-occurrence* (janvier 2005)

n°19 Jacques FRANÇOIS*, Les bases épistémologiques des grammaires cognitives et 'néo-fonctionnelles'* (juin 2005)

n°20 Jacques FRANÇOIS*, Le fléchage synonymique de la polysémie verbale* (octobre 2005)

n°21 Dominique LEGALLOIS & Jacques FRANÇOIS (coordination), *Autour des grammaires de constructions et de patterns* (janvier 2006) [avec la participation de Philippe GREA et Morgane SENECHAL]

n°22 Odile BLANVILLAIN*, Le marqueur SUCH en anglais contemporain – Anaphore qualitative et identification notionnelle* (mai 2006)

n°23 Jacques FRANÇOIS & Ahmed BRAHIM (coordination)*, Morphosyntaxe et sémantique du verbe* (novembre 2007)

n°24 Jacques FRANÇOIS*, Une approche diachronique quantitative de la polysémie verbale* (janvier 2008)

n°25 Mayu SHINTANI*, Une vision rhétorique des expressions BOUT et HASHI – universalité cognitive et comparaison entre français et japonais* (février 2008)

n°26 Jacques FRANÇOIS, *Les grammaires de construction, un bâtiment ouvert aux quatre vents* (septembre 2008)

n°27 Lise HAMELIN, *L'alternance entre construction ditransitive et complémentation prépositionnelle : étude des marqueurs* ***to*** *et* ***for*** (mai 2009)

n°28 Franck NEVEU (coordination), *Travaux des doctorants du CRISCO 2008-2009* (octobre 2009)

n°29 Jacques FRANCOIS, *L'attestation des combinaisons lexicales à l'aide de la base de données FRANTEXT* (janvier 2010)

n°30 Lise HAMELIN (coordination), *Travaux des doctorants du CRISCO 2009-2010 (mai 2010)*

n°31 Charlotte BAILLEUL (coordination), *Travaux des doctorants du CRISCO 2010-2011 (juin 2011)*

n°32 Gaëlle DOUALAN, *Introduction à une approche instrumentée de la synonymie (septembre 2011)*

n°33 Charlotte BAILLEUL (coordination), *Travaux des doctorants du CRISCO 2010-2011* (novembre 2011)